# NOTTINGHAM LAW JOURNAL

## VOLUME 14

## 2005

# NOTTINGHAM LAW JOURNAL

VOL 14(1) 2005

## CONTENTS

NOTTINGHAM LAW JOURNAL

VOL 14(1) 2005

# EDITORIAL

It is in the nature of any sensible and flourishing enterprise that it should, like the Roman god Janus, constantly look backward and forward at the same time. We need to look backward, not for reasons of self-indulgence, but to be clear about our traditions, our values, our strengths; in other words, the very things that helped to make us what we are. We need to look forward and plan for the future, but always with an eye to building on the firm foundations that have been laid.

It is in this spirit of looking backward and forward that we at the *Nottingham Law Journal* wish to record our thanks and best wishes to Professor Peter Kunzlik, who has recently moved from Nottingham Law School to take up the post of Director of the City Law Institute and Pro-Vice-Chancellor of the City University. In the limited space available to me it is not possible to do full justice to the significant contribution that Peter has made to the life of our Law School. But one thing, as editor, I must say is that there is no doubt that Peter's vision and energy have been crucial to the development and sustaining of this journal. So we look backward in celebration of the contribution that Peter has made and we look forward to him remaining as a member of our Advisory Board and continuing his association with us!

ADRIAN WALTERS

NOTTINGHAM LAW JOURNAL

VOL 14(1) 2005

# ARTICLES

## THE PROBLEM OF *CLARENCE*, BRUTAL AT HIS BEST

## JAMES FITZJAMES STEPHEN AND THE DOCTRINE OF SEXUAL INEQUALITY

*KATE GLEESON**

### PROSECUTING DISEASE TRANSMISSION

In October 2003 Mohammed Dica was sentenced to eight years for inflicting grievous bodily harm ("GBH") by recklessly passing on HIV to two women with whom he had sex.[1] In January 2004, Kouassi Adaye pleaded guilty to inflicting GBH by transmitting HIV to one woman and was sentenced to six years.[2] In May 2004 Feston Konzani was sentenced to ten years for causing GBH by transmitting HIV to three women.[3] These charges and convictions are controversial because they seem to run against common law precedent for understanding assault by sexually transmitted disease.[4] The leading case in this area is the 1888 case of *Clarence,* in which it was ruled that the deceitful transmission of gonorrhoea did not constitute an assault.[5] As Charles Foster notes, "*Clarence*, rooted as it is in a Victorian morality, has ruled this arcane but important corner of the law ever since".[6] The appeal court in *Dica* seems now to have overruled *Clarence*, although the case is pending further appeal.[7]

His Honour Judge Philpot, the trial judge in *Dica,* argued that the standing of *Clarence* "as an important precedent has been thoroughly undermined . . . and provides

* PhD candidate, School of Politics and International Relations, University of NSW. Research for this article was undertaken with the SOLON research project of the partnership between Nottingham Trent University and the Museum of Law. My participation in the SOLON project was supported by a British Council Chevening Scholarship and an Australian Universities European Network travel grant, for both of which I am grateful. Thanks also to Nottingham Law School, Dr Kim Stevenson and Dr Judith Rowbotham of NTU for their support and advice.

[1] Charles Foster, "The Questions to ask a Husband (or Wife) Before Sex", *Times* (London), 21 October 2003, 6.

[2] BBC News, 12 January, 2004, "HIV Bigamist Jailed for Infections": http://news.bbc.co.uk.

[3] Michael Taggart, "Jailed for Ten years – the One-man HIV Epidemic", *Daily Mail* (London), 15 May 2004, 29.

[4] The convictions are also controversial as the three men are all asylum seekers from Africa. This raises important considerations that I do not address here.

[5] *R v Clarence* (1888) 16 Cox CC 511.

[6] Foster, *op cit*, 6.

[7] Dica's appeal commenced at the Inner London Crown Court on 21 June 2004. On 23 June 2004 the jury was discharged "for legal reasons". A date has not been set for the new hearing (BBC News, 24 June, 2004, "Fresh Trial for HIV Accused Man": http://news.bbc.co.uk).

no guidance to a [first] instance judge".[8] On 5 May 2004 Dica's appeal was allowed on the basis that Philpot J should not have withdrawn the issue of consent from the jury.[9] On appeal *Clarence* was again dismissed as of "no continuing relevance" and as no longer "authoritative" on the question of consent to sex with risk of disease.[10] While *Clarence* appears to have been dismissed as authority, this is not the first time it has been disputed. *Clarence* is most well known for its notorious affirmation of the marital rape exemption. Yet since the House of Lords ruled that the exemption was no longer good law in 1992,[11] *Clarence* has continued to be cited in a number of cases, not least in defining the boundaries of consent.[12] *Clarence* is a tenacious case that has somehow survived the overruling of its anachronistic logic.

Despite its "Victorian morality", *Clarence* is the modern authority on consent within sex and beyond. It continues to be cited as persuasive authority in cases ranging from rape to sadomasochism, and even to negligent drunken dentistry.[13] Indeed, despite *Dica,* Carol Withey suggests that *Clarence* yet again "may see another day".[14] The recent HIV convictions have brought *Clarence* once more into the spotlight. The judgment more than ever seems ripe for review, yet it is a complicated judgment the authority of which, in the complex and evolving area of consent and sex, has not been analysed in depth. Despite being notorious, *Clarence* is generally poorly understood.

In this article I examine *Clarence* in detail. I argue that the contemporary use of *Clarence* is peculiar in any area, as the judgment is steeped in a particular morality and archaic prejudices concerning women's rights and the criminal law's interest in sexual harm. I explore how the judgment was formed, and consider its underlying motivation and intention. I identify the agenda of the bench discernible within the decision and explain the prevailing moral and political climate that informed the judgment. In particular, I focus on the role of Sir James Fitzjames Stephen and the prejudices that he brought to bear in formulating the judgment. I highlight Stephen's personal antipathy to equality, particularly within marriage, and his persistent refusal to acknowledge harm in sex outside the most physically violent of circumstances, as well as his callous approach to recognising sexual fraud.

At the time and for some years after, *Clarence* was a fragile settlement. It was a controversial and politically motivated decision, and an anomaly in its denial of women's rights even in 1888. It stood in stark contradiction to the prevailing trend towards the recognition of sexual equality in the courts and parliament. The decision was in ideological opposition even to the prevailing Victorian ideas about women's rights. Its fragility and its inconsistency with contemporary values makes its endurance seem all the more remarkable – and all the more inappropriate. Examining *Clarence*, I conclude that the fundamental limitations of the judgment cannot be overcome and that it should not be considered an authority on the nature of consent in any field. Despite my criticisms of *Clarence*, however, I would not argue for the prosecution of

[8] Cited in *R v Dica* [2004] EWCA Crim 1103, [2004] QB 1257.

[9] Philpot J concluded that whether or not the complainants knew of the appellants' condition, their consent, if any, was irrelevant and provided no defence. Accepting the Crown's argument as advanced to him, Philpot J believed that the decision in the House of Lords in *R v Brown & ors* [1994] 1 AC 212 deprived the complainants of "the legal capacity to consent to such serious harm" (in *R v Dica* [2004] QB 1257).

[10] *R v Dica* [2004] QB 1257.

[11] *R v R* [1992] 1 AC 599.

[12] *R v Brown* [1994] 1 AC 212; *R v Linekar* [1995] QB 250, [1995] 3 All ER 68; *R v Burstow* [1998] AC 147 *R v Cuerrier* (1997) 111 CCC (3d) 261 (CACB); *R v Richardson* [1999] QB 444.

[13] *R v Linekar* [1995] 3 All ER 69; *R v Brown* [1994] 1 AC 212; *R v Richardson* [1999] QB 444.

[14] Carol Withey, "Biological GBH: Overruling Clarence?", (2003) 153(7104) *New Law Journal*, 698.

GBH for HIV transmission along the lines of *Dica*. This area of criminal law requires urgent and critical review, separate from the authority of *Clarence,* with the development of thinking that incorporates, as Matthew Weait notes, "alternative interpretations of harm".[15]

## CHARLES AND SELINA CLARENCE

In 1888 Charles Clarence was charged with assault occasioning actual bodily harm and inflicting grievous bodily harm, for his knowing or reckless transmission of gonorrhoea to his unwitting wife. He was convicted of both counts of assault at the Central Criminal Court. The Recorder of London directed the jury that, "if the evidence established the facts hereinbefore set out to their satisfaction then, notwithstanding the fact that Selina Clarence is the prisoner's wife, they might find him guilty on both counts, or either count of the indictment".[16] The jury "strongly recommended mercy"[17] in sentencing and the Recorder was wary of the convictions based on his direction. He referred the case for further clarification to the Queen's Bench, with the request, "I desire the opinion of the court, whether upon such direction and facts the prisoner could properly be convicted on both counts, or either count of the indictment".[18] Clarence was released on £150 surety, paid by his mother.[19]

The case was initially referred to a sitting of five judges of the Queen's Bench, who quickly realised their "considerable differences of opinion" about the case.[20] Lord Coleridge CJ noted that the case was one "of the greatest importance, involving questions of serious difficulty".[21] He explained that the case would be held over until the defendant could be provided counsel by Treasury, and until all 15 judges of the Queen's Bench were available to hear the "important" case.[22] In fact, 13 of the bench presided, with one of the 15 "being abroad on account of ill-health" and the other "being at Chambers".[23] Initially Mr Poland for the prosecution had explained that "knowing that the prisoner was not represented by counsel, he had felt it to be his duty, besides preparing his own argument for the Crown, to examine the law and decisions which might be deemed in favour of the accused also".[24] However his efforts were not necessary as the Treasury appointed Mr Forrest Fulton to defend Charles Clarence.

Clarence had been convicted of assault and GBH, offences not involving a constitutive component of consent. However the Queen's Bench case was creatively introduced as concerning the "marital contract" of *consent to sex*. Much debate involved the questions of whether Selina had consented to sex with her husband, and whether her consent was vitiated by the fact that she was unaware of his disease. Defence counsel, Forrest Fulton, introduced Clarence's case with a manipulation of the concept of "unlawful". Fulton argued that as outlined in the Offences Against the

[15] Matthew Weait, "Taking the Blame: Criminal Law, Social Responsibility and the Sexual Transmission of HIV", (2001) 23(4) *Journal of Social Welfare and Family Law,* 441.

[16] Noted in *R v Clarence* (1888) 16 Cox CC 511 at 512.

[17] PRO Crim 8/1.

[18] Noted in *R v Clarence* (1888) 16 Cox CC 511 at 512.

[19] PRO Crim 4/1030 118901.

[20] *Times* (London), 11 June 1888, 4.

[21] *Ibid.*

[22] *Ibid.*

[23] *Times* (London) 18 June 1888, 4.

[24] *Times* (London) 11 June 1888, 4.

Person Act 1861, the offence of GBH involved the "unlawful" infliction of harm. Therefore, he argued, it was necessary to establish both "that the act of the prisoner was unlawful, and that grievous bodily harm ensued in consequence of such an unlawful act".[25] While Fulton conceded that grievous bodily harm followed from Clarence's actions, he creatively claimed that "the act which commissioned it was not unlawful".[26] By this logic, GBH is an offence only when inflicted unlawfully.[27] This interpretation suggests that the "harm" with which the law is concerned is not an objective category identified by its measurable effects on victims. Instead, harm gains its meaning and character first and primarily through delineation by the law. Therefore some forms of recognizable, perhaps even lamentable, bodily harm, must by definition be considered lawful. This distinction relates neither to the *measure* nor the *experience* of harm (nor to the intention of the infliction). This view involves a particular understanding of the role of law as well as the nature of harm. By this logic, the object of the criminal law of assault is not primarily to prevent or to redress harm. Rather, its role is to demarcate which, or whose, experience of harm is unlawful.

In *Clarence*, "consent" was argued to determine the difference between lawful and unlawful bodily harm. The Offences Against the Persons Act seemed potentially to allow for "lawful" infliction of GBH, and Fulton submitted that Selina's consent to the sex by which she had been infected had determined that Clarence's actions were not unlawful. Selina had "consented" to sex with Charles, therefore his actions were not unlawful, and therefore the ensuing GBH was not unlawful. With regard to Clarence's second conviction for assault occasioning actual bodily harm, Fulton argued that in order to support a conviction "it is necessary to prove an assault, and an assault implies an absence of consent".[28] The offence of assault occasioning actual bodily harm does not contain the magic transformative word "unlawful", on which Fulton's arguments against the GBH conviction depended. Therefore Fulton inferred a constitutive component of "consent" for the charge of assault.[29]

## IRREVOCABLE CONSENT

In support of the argument that Selina had consented, the 17th century pleas of Sir Matthew Hale were introduced. Fulton argued that there was "high authority" for deciding that the "prisoner's act here was lawful".[30] According to Lord Hale, "the husband cannot be guilty of a rape committed by himself upon his lawful wife, for by their mutual matrimonial consent and contract the wife hath given herself up in this kind unto her husband which she cannot retract".[31] This made the question of disputed consent irrelevant in the context of marriage. Fulton argued that this was "law in Lord Hale's time and is still the law",[32] despite the fact that Hale's pronouncement seemed not to have been tested or affirmed once in the 150 years since its publication. Thus, in an attempt to identify a not unlawful form of GBH, Hale's dictum on irrevocable

[25] *R v Clarence* (1888) 16 Cox CC 511 at 512.

[26] *Ibid.*

[27] In *Brown* the House of Lords clarified that "not unlawful" pertains to instances of self-defence ([1994] 1 AC 212).

[28] *R v Clarence* (1888) 16 Cox CC 511 at 512.

[29] In *Brown* the House of Lords clarified that consent is not relevant to assault occasioning actual bodily harm (*R v Brown* [1994] 1 AC 212).

[30] *R v Clarence* (1888) 16 Cox CC 511 at 512.

[31] Sir Matthew Hale, "Concerning Felonies by Act of Parliament and first Concerning Rape", in GR Glazebrook (ed), *The History of the Pleas of the Crown Volume One*, (1736), (Professional Books Limited, London, 1971) 629.

[32] *R v Clarence* (1888) 16 Cox CC 511 at 512.

marital consent was introduced to the Queen's Bench. Fulton advised in his opinion, of the perils of denying Hale's dictum as a result of the argument that Selina's consent was vitiated because she was not aware of her husband's disease. He warned that the consequence of departing from

> the principle that the wife's consent is irrevocable on the ground that fraud vitiates her consent would be that any kind of fraud would vitiate such consent; and any deceit practised by a person by means of which he obtained a woman's consent would render such person guilty of rape.[33]

Fulton's argument was first, that Clarence's *act* of inflicting GBH was not unlawful, because of the doctrine of irrevocable consent. And second, that it *must* be considered lawful, because of the legally impractical ramifications that would ensue from denying irrevocable consent.

Hale's opinion on marital rape was written in around 1670 and later published in his *History of the Pleas of the Crown*. However, it is unclear on what authority Hale based his decree. There was "no similar statement in the works of any earlier English commentator",[34] and Hale provided no authority for his pronouncement, which was unusual in his work. As Colin Manchester, *et al*, note,

> It was not that the custom of offering authority for points of law was undeveloped in the 17th century. Indeed on the two pages containing the passage on marital rape there are seven references to cases and statutes as authority for statements of the law on other points related to rape.[35]

Hale's work was unfinished when he died in 1676. He had not authorised for publication the *History of the Pleas of the Crown,* and a special resolution of the House of Commons was required for it to be published in 1736.[36] Bruce McFarlane argues that it is unclear whether Hale intended the widespread publication of his treatise, wholly unrevised as it was.[37] The idea that on marriage a wife gave her body to her husband was accepted in matrimonial cases decided in the Ecclesiastical Courts, but with limitations. In the 1794 civil case of *Popkin v Popkin,* Lord Stowell directed that "the husband has a right to the person of his wife, *but not if her health is endangered*".[38] For 150 years after Hale's direction, there appears to have been no reported case that invoked the ruling in the criminal law. It remained unconfirmed in the criminal courts until *Clarence,* where Fulton's arguments were persuasive. It was ruled that Selina Clarence had consented to the sex, there had been no unlawful assault, and Clarence's convictions were to be quashed. The ruling promoted the "offensive fiction" that a wife's consent to sex with her husband was irrevocable and could not be vitiated in any circumstances.[39] This is despite the fact that the original trial convictions did not concern rape, but rather assault and GBH.

[33] *Ibid.*

[34] *Ibid.*

[35] Colin Manchester *et al.*, *Exploring the Law: The Dynamics of Precedent and Statutory Interpretation,* (Sweet & Maxwell, London, 1996) 310. In my doctoral thesis I explore the limitations and inconsistencies of Hale's *dicta* on rape.

[36] Bruce McFarlane, "Historical Development of the Offence of Rape", *Canadian Criminal Law*, available at http://canadiancriminallaw.com/articles/abstracts/hist_rape.htm (1993) 35.

[37] *Ibid.*

[38] *Popkin v Popkin* (1794) 1 Hag Ecc 765n, *per* Lord Stowell, cited in *R v R* [1992] 1 AC 599, *per* Lord Lane, 604. Emphasis added.

[39] *R v R* [1991] 1 All ER 747, *per* Owen J, 749.

## A LACK OF UNANIMITY

After *Clarence*, the common law "marital rape exemption" was relied upon for over 100 years to protect husbands from charges of rape for forcing sex on their wives. Some have argued that in invoking the exemption, *Clarence* simply affirmed the prevailing Victorian common law understanding of a wife's rights. For example, Damian Warburton writes that "in 1888 a wife had no right to withhold consent and her husband could not be found guilty of raping her".[40] However, this analysis ignores the radical, normative and stipulative nature of the *Clarence* decision and of the judgment of Stephen J in particular. Prior to *Clarence*, the view that a wife's consent was irrevocable had not been affirmed in the criminal law. Even within the majority of *Clarence* there was no unanimity regarding the grounds of the opinion. In fact most of the judges in *Clarence* did not accept Hale's pronouncement absolutely.

Wills J, who gave the first judgment and with whom Lord Coleridge CJ concurred, stated that he was "certainly not prepared to assent" to the proposition that rape in marriage was impossible.[41] For Wills J, there seemed to be "no sufficient authority" for the proposition at all, and it was "equally true that a married woman, no less an unmarried woman, would be justified" in refusing sex, had her husband disclosed his disease to her.[42] Yet still he held that to affirm the convictions would create untenable implications. Wills J agreed with the paranoid logic of the defence that the convictions might mean that "a wide door will be opened to inquiries not of a wholesome kind", as where a "man meets a woman on the street and knowingly gives her bad money in order to procure her consent to intercourse with him".[43] In such an instance, for Wills J, "it would be childish to say she did not consent".[44]

The chastisement of Wills J of "childish" complaints of sexual harm suggests he was unwilling to conceive of sex in terms of a legitimate or "real" harm, despite women's experiences and complaints to the contrary. Wills J was mainly interested in cases of prostitution and "seduction" and he was concerned with the difficulties he foresaw in "arriving at the truth" in cases of disputed consent.[45] He ruled that rape in marriage was possible, but that in the case of Mrs Clarence her consent was real and legitimate. Thus, he held, there was no assault.[46]

Hawkins J also disputed the extent of Hale's authority. In what has been described by Prowse J as a "spirited dissent",[47] he argued that the convictions should be upheld and questioned whether Hale's opinion should be applied in the case of physical harm. While Hawkins J agreed that the marital contract "conferred upon the husband the irrevocable privilege to have sexual intercourse with her" in the case of "normal relations", he argued that "this marital privilege does not justify a husband in

[40] Damian Warburton, "A Critical Review of English Law in Respect of Criminalising Blameworthy Behaviour by HIV+ Individuals", (2004) 68 *Journal of Criminal Law*, 63.

[41] *R v Clarence* (1888) 16 Cox CC 511 *per* Wills J, 520.

[42] *Ibid.*

[43] *Ibid.*

[44] *Ibid.*

[45] *Ibid.*

[46] In *Coney*, a mere six years prior to *Clarence*, it had been clarified that consent is all but irrelevant to the crime of assault. *Coney* involved the appeal of three men for their convictions of aiding and abetting common assaults after they were arrested for watching a prizefight. The convictions were quashed with the ruling that while the prizefight did constitute common assaults, the men's presence in the audience could not be proven to constitute aiding and abetting the assaults. In clarifying the illegal nature of prizefighting, and its connection to assault, Cave J explained the irrelevance of consent as follows: "the true view is, I think ... that a blow struck in anger, or which is likely or is intended to do corporal hurt, is an assault ... and that an assault being a breach of the peace and unlawful, *the consent of the person struck is immaterial*" (*Queen v Coney and Others* (1882) LR 8 QBD 535, *per* Cave J, 539. Emphasis added). *Coney* was not cited in *Clarence*. In 1993 this view was confirmed by the House of Lords. See n 29 above.

[47] *R v Cuerrier* (1997) 111 CCC (3d) 261 (CACB), *per* Prowse J.

endangering his wife's health and causing her grievous bodily harm".[48] Field J, with whom Charles J concurred, also argued for the convictions to be upheld. Field J had presided over the 1877 trial of *Flattery* that first recognised rape by medical fraud, a ruling that Stephen J criticised as incorrect.[49] Field J was also critical of the use of Hale's opinion, stating,

> the authority of Hale C J on such a matter is undoubtedly as high as any can be, but no other authority is cited by him for his proposition, and I should hesitate before I adopted it. There may, I think, be many cases in which a wife may lawfully refuse intercourse, and in which, if the husband imposed it by violence, he might be guilty of a crime. Suppose a wife for reasons of health refused to consent to intercourse, and the husband induced a third person to assist him while he forcibly perpetrated the act, would anyone say that the matrimonial consent would render this no crime?[50]

It is not clear how the marital rape exemption came to be considered so orthodox a rule throughout the 20th century, given the lack of unanimity and the variety of opinions of the bench in *Clarence* about the exemption.

In contrast to the dissenting opinions was the less sympathetic view of Pollock CB, who seems to have been quite an admirer of Lord Hale. Pollock CB argued that:

> the wife as to the connection itself is in a different position from any other woman [and] such a connection may be accompanied with conduct which amounts to cruelty, as where the condition of the wife is such that she will or may suffer from such connection".[51]

However, Stephen J's opinion attracted most support, and is considered by Bruce McFarlane and others to have "formed the decision of the court".[52] Stephen J's opinion was supported explicitly by four of the 13 judges in the case. He focussed on the nature of disease and assault as it was understood from the legislation, and also on the implications of acknowledging that fraud vitiates consent to sex.

The arguments used by Stephen J in support of quashing Clarence's convictions are sometimes explained as resting on his direction concerning the nature of assault and the inflicting of GBH as defined by statute. Stephen J argued:

> the words "infliction of bodily harm" . . . appear to me to mean the direct causing of some grievous injury to the body itself with a weapon, as by a cut with a knife, or without a weapon, as by a blow with a fist or by pushing a person down. Indeed, though the word "assault" is not used in this section, I think the words imply an assault and battery of which a wound or grievous bodily harm is the manifest immediate and obvious result.[53]

The point is developed by Damian Warburton, who concludes that:

> the majority in *Clarence* could not reconcile what they saw as a requirement for an immediate, violent, and obvious harm resulting in a wound, as they saw the word "inflict" within the statute to mean, with infection resulting from consensual intercourse.[54]

[48] *R v Clarence* (1888) 16 Cox CC 511, *per* Hawkins, J, 532.

[49] *R v Flattery* (1876–1877) LR 2 QB 410. Stephen J had criticised the decision in *Flattery*, asserting "if such permission is given, the fact that it was obtained by fraud, or that the woman did not understand the nature of the act, is immaterial" (*Digest of Criminal Law (Crimes and Punishments)*. 3rd edition, Macmillan & Co., London, 1883, 185–86).

[50] *R v Clarence* (1888) 16 Cox CC 511, *per* Field J, 537–38.

[51] *Ibid*, *per* Pollock CB, 542.

[52] McFarlane, *op cit*, 32.

[53] *R v Clarence* (1888) 16 Cox CC 511 *per* Stephen J, 525.

[54] Warburton, *op cit*, 60.

This is a common interpretation of the case. Again, for David Ormerod, "a husband's knowing transmission of the disease to his wife by acts of sexual intercourse was held not to constitute an infliction because there was no assault".[55]

However, this was not the primary concern of Stephen J, as he clarified in the 1889 case of *Moss*.[56] The case of *Moss* concerned a married couple's convictions for GBH for child abuse, primarily by starving their child. Wills J had referred the case to the Queen's Bench for clarification. Wills J was concerned that as the abuse resulted from starvation and not physical assault, he should perhaps have directed an acquittal in light of *Clarence*. He explained that in *Clarence* "some of the Judges of this Court were supposed to have held, as to a charge of inflicting bodily harm, that there must be some act of violence".[57] Lord Coleridge CJ noted that "in the case of *Clarence* Mr Justice Stephen held that 'inflict' means some act of violence".[58] However, Stephen J clarified his position explicitly and curtly in *Moss*, stating: "that was a very subordinate part of my judgment, and I rather protested against philological discussions".[59]

Like Wills J, Stephen J in *Clarence* was preoccupied with the implications of the case for sex outside marriage, especially cases of "seduction" and prostitution and "promises not intended to be filled": cases where women might seek recourse for the harm they identified in sex.[60] Stephen J referred to Hale's irrevocable consent doctrine in *Clarence* only in his closing remarks to clarify that in the latest (then forthcoming) edition of his *Digest of the Criminal Law* he had withdrawn the statement in earlier editions that a husband might, under certain circumstances, be indicted for the rape of his wife.[61] In the 1877 edition of the *Digest,* Stephen J had written:

> Hale's reason is that the wife's consent at marriage is irrevocable. Surely, however, the consent is confined to the decent and proper use of marital rights. If a man used violence to his wife under circumstances in which decency or her own health or safety required or justified her in refusing her consent, I think he might be convicted of rape, notwithstanding Lord Hale's dictum.[62]

While it is not clear why he had changed his mind about marital rape in the years before *Clarence,* Stephen J claimed to have come wholly to support Hale's opinion.[63] But, even so, the judgment of Stephen J was representative of only five of the 13 on the bench. If one includes Pollock CB's mandate for cruelty, only six of the bench followed Hale's opinion. The creation of the marital rape exemption was far from straightforward or unanimous. This observation has led Jocelynne Scutt to comment

[55] David Ormerod, "Criminalising HIV Transmission – Still no Effective Solutions", (2001) 30(2) *Common Law World Review*, 138. In 1998 it was clarified that "inflict" could include psychological harm (*R v Burstow* [1998] AC 147).

[56] *R v Anne Moss and William Thomas Moss* (1889) TLR 224.

[57] *Ibid, per* Wills J.

[58] *Ibid, per* Lord Coleridge CJ, 225.

[59] *Ibid, per* Stephen J.

[60] *R v Clarence* (1888) 16 Cox CC 511, *per* Stephen J, 527.

[61] *Ibid.*

[62] James Fitzjames Stephen, *A Digest of Criminal Law (Crimes and Punishments),* Fourth Edition, (Macmillan & Co, London, 1887) 186.

[63] In fact the paragraph was not omitted. It was altered. The matter is confused by Stephen's death. Stephen died in March 1894 before the fifth edition of his *Digest* was published later that year. The fifth edition is edited by his sons Sir Herbert Stephen and Harry Lushington Stephen who professed to have "endeavoured to make it as nearly as possible what it would have been if it had been prepared by the author" (Stephen and Stephen (eds), *A Digest of the Criminal Law (Crimes and Punishments) by the Late Sir James Fitzjames Stephen, Bart*, fifth Edition, MacMillan & Co, London, 1894, preface). In this edition Stephen's commentary on Hale's irrevocable consent doctrine suggests that a husband might be found guilty of *indecent assault* of his wife, rather than rape, as the earlier editions had directed. The fifth edition states: "[i]f a man used violence to his wife under circumstances in which decency or her own health or safety required or justified her in refusing her consent, I think he might be convicted at least of an *indecent assault*" (208. Emphasis added). It is unclear whether Stephen or his sons authorised the alteration.

that despite popular legal mythology sustained throughout the following century, *Clarence* was "not the bulwark of the irrevocable consent argument it may have been considered".[64]

Even at the time, it is doubtful whether *Clarence* was understood to have confirmed the marital rape exemption, or to have been a particularly important judgment. In 1894, the fifth edition of Stephen J's *Digest of the Criminal Law* was edited posthumously by his sons Sir Henry Stephen and Harry Lushington Stephen. They explained in the preface that

> not many cases of the first importance have been decided since the publication of the Fourth Edition at the beginning of 1887 ... All cases which seem either to alter or more fully to expound the law, have been incorporated either into the text or by way of illustration.[65]

*Clarence* is not cited in the *Digest*. It is not clear from whence its significance and authority is understood first to have arisen. It seems that in the 60 years after *Clarence*, marital consent was not disputed within the criminal law. By the time it came to be tested in 1949, the mythology of the marital rape exemption had been cemented as indisputable in the common law.[66]

## A WIFE'S RIGHTS – JACKSON V CLARENCE

As I indicated above, it could perhaps be said that a judgment, the effect of which was to confirm women's subordinate status in marriage and their sexual liberties as contingent on men, merely reflected the political and legal climate in which women resided in late Victorian England. I have argued, however, that *Clarence* was a far from unanimous judgment, and it was, in fact, delivered in a period of a dynamic and contested political and legal discourse over the status of women. The conclusion of those judges of the Queen's Bench in *Clarence* who affirmed Hale's 17th century opinion need not have been assured or predictable. In the same period, the Victorian bench was vigorously engaging with and disputing early legal authority in fields similar to those discussed in *Clarence*.

Perhaps the most striking case was the 1891 case of *Jackson,* in which early legal opinions on marriage by authors such as Bacon and Hale were introduced by defence counsel, disputed vehemently and finally dismissed by the bench. *Jackson* concerned the defendant's "imprisonment" of his wife, and her petition for a writ of *habeas corpus*.[67] Emma Jackson had complained of the deprivation of her liberties involved in her husband's detaining of her in their house in order to "enforce restitution of [his] conjugal rights".[68] Edmund Jackson's behaviour was defended by counsel on the basis that "a husband has a right to take the person of his wife by force and keep her in confinement, in order to prevent her from absenting herself from him so as to deprive him of her society".[69] The defence arguments were supported with reference to "early

[64] Jocelynne A Scutt, "Consent in Rape: The Problem of the Marriage Contract", (1977) 3 *Monash Law Review*, 260.

[65] Stephen and Stephen (eds), *A Digest of the Criminal Law (Crimes and Punishments)*, 1894, *op cit*, preface.

[66] *R v Clarke* [1949] 2 All ER 448.

[67] *R v Jackson* [1891] 1 QB 673.

[68] *Ibid.*

[69] *Ibid, per* Lord Esher.

authorities" of the opinion that a husband "is entitled to the custody and control of his wife, and to detain her by force if she refuses to live with him", and included reference to Hale's irrevocable consent doctrine.[70]

However in this instance the bench did not accept the "early" version of marriage, and harshly criticised the invocation of anachronistic legal arguments. In Lord Halsbury's opinion, the defence was advocating slavery. He argued,

> More than a century ago it was boldly contended that slavery existed in England, but if anyone were to set up such a contention now, it would be regarded as ridiculous. In the same way the quaint and absurd *dicta* as are to be found in the books as to the right of a husband over his wife in respect of personal chastisement are not, I think, capable of being cited as authorities in a court of justice in this or any civilised country.
> It is important to bear this in mind, for many of the statements which have been relied upon of a more moderate character and less outrageous to common feelings of humanity, are bound up with these ancient *dicta* to which I refer ... It appears to me that the authorities cited for the husband were all tainted with his sort of notion of the absolute dominion of the husband and wife.[71]

Hale's 17th century ideas on irrevocable consent might generally have been considered as of a "more moderate" and "less outrageous" character in 1891, than, say, the doctrine of lawful marital chastisement or the right of a husband to detain his wife argued for in *Jackson*. However, for Lord Halsbury, even such "moderate" views were tainted with the uncivilised notion of the "absolute dominion of the husband and wife" that he disputed as absurd.[72] For Lord Halsbury, the rejection of one part of the early authorities' ideas on marital rights cast doubts on the appropriateness of any other part of their doctrine. He asked "where ancient *dicta*, which state that a husband is entitled to imprison his wife, also state that he has a right to beat her, can they be rejected as authorities for the latter proposition, without being affected as authorities for the former?"[73] Lord Halsbury's insightful questions remain relevant today. It is questionable whether the *Clarence* judgment, which was responsible for the denial of the harm of sex within marriage, is an appropriate legal authority in terms of sex and consent at all. Can we extract sense and relevance from *Clarence* even after its logic, in the form of the marital rape exemption, was rejected as "offensive fiction"?[74] On Lord Halsbury's logic, it is at best questionable.

Hale's *dictum* on irrevocable consent was not explicitly disputed in *Jackson*. However, implicit in the debates of the court, was the *object* of detainment: the marital right to sex. This suggests that sex as the object of the detainment was understood not to be a compelling factor in understanding the husband's rights of marriage. Detainment was ruled unlawful, despite or regardless of its purpose. This view could be understood to diminish, or at least to complicate, the practical realisation of Hale's doctrine; for it suggests that a husband may not detain his wife in order to have her fulfil the "marital obligation" that had been upheld as irrevocable by the courts only three years earlier. The rulings of *Clarence* and *Jackson* are difficult to reconcile. In combination, they suggest that a woman is obliged to have sex with her husband as part of her marital duty, but that, at the same time, her husband may not lawfully compel her to fulfil her duty by means of detainment or chastisement. On this basis, it is difficult to see how a husband might legally have realised his "right" to marital

[70] *Ibid*, 675.
[71] *Ibid*, *per* Lord Halsbury, 678–79.
[72] *Ibid*, *per* Lord Halsbury, 679.
[73] *Ibid*, *per* Lord Halsbury, 675.
[74] *R v R* [1991] 1 All ER 747, *per* Owen J, 749.

sex should his wife have refused him. In fact Hale's own views on lawful chastisement seem also to suggest a contradiction. Lord Halsbury explained that Hale, in his "free translation" of early doctrine, had suggested that "*castigatio* may be taken to mean admonition merely".[75] It seems that Hale did not advocate or uphold the beating of wayward wives, though imprisonment in the form of "confinement to the house in the case of the wife's extravagance" was acceptable to him.[76] Again, it is unclear how in this view it was expected that a husband might have exercised his right to marital sex, should the woman have resisted.

Despite the apparent misogyny and justifications of marital cruelty in *Clarence*, no judge advocated the use of the law to compel a woman to fulfil her "duty". In those judgments supporting Hale's opinion, it was argued that a wife could not revoke her consent, but generally the issue of how the "obligation" might be legally enforced was avoided. The tension apparent between the "right" of marital sex and the "right" of a wife's physical liberty that must be contested in this combined common law view of marriage illuminates the complex and evolving discourse of equality that was emerging in England and Wales in the 19th century. It illustrates moreover how prejudices and personalities on the bench influenced the interpretation and creation of the common law. The assessment of Hale's opinion on irrevocable consent could in principle have fallen either way in this climate. Perhaps Lord Halsbury's verdict would have been different from the majority in *Clarence*, had he presided over the case. In *Jackson*, Lord Halsbury dismissed Hale's opinion on a husband's right to confine his wife as absurd dicta, and argued that such a dismissal should suggest that all similar *dicta* from the early authors on marriage no longer could be supported. In 1891 Lord Halsbury demanded emphatically and nobly of Emma Jackson that "this lady must be restored to her liberty".[77] What might he have made of the marital rape exemption, had he presided over *Clarence* in 1888?

## EMERGING EQUALITY

Lord Halsbury's rhetoric in *Jackson* of progress and evolution from "quaint" and "absurd" dicta was in keeping with the evolving spirit of sexual equality of the late 1800s that dwelt on notions of "transcended barbarism".[78] John Stuart Mill had published his radical treatise *The Subjection of Women* in 1869, in which he argued that "the principle which regulates the existing social relations between the two sexes – the legal subordination of one sex to the other – is wrong in itself, and now one of the chief hindrances to human improvement".[79] For Mill, marriage was akin to slavery in its oppression of women's rights. He wrote,

> I am far from pretending that wives are in general no better treated than slaves; but no slave is a slave to the same lengths, and in so full a sense of the word, as a wife is. Hardly any slave, except one immediately attached to the master's person, is a slave at all hours

[75] *R v Jackson* [1891] 1 QB 673, *per* Lord Halsbury, 675.

[76] *Ibid.*

[77] *Ibid,* 682.

[78] John Stuart Mill warned that we do not "suppose that the barbarisms to which men cling longest must be less barbarisms than those which they earlier shook off" (John Stuart Mill, "The Subjection of Women" (1869), *On Liberty and Other Essays,* edited by John Grey, Oxford University Press, Oxford, 1992, 471–472).

[79] *Ibid,* 471.

> and all minutes; in general he has, like a soldier, his fixed task, and when it is done, or when he is off duty, he disposes, within certain limits, of his own time, and has a family life into which the master rarely intrudes.[80]

For Mill, a husband's cruelty had impunity within marriage:

> "Uncle Tom" under his first master had his own life in his "cabin", almost as much as any man whose work takes him away from home, is able to have in his own family. But it cannot be so with the wife. Above all, a female slave has (in Christian countries) an admitted right, and is considered under a moral obligation, to refuse to her master the last familiarity. Not so the wife: however brutal a tyrant she may unfortunately be chained to — though she may know that he hates her, though it may be his daily pleasure to torture her, and though she may feel it impossible not to loathe him — he can claim from her and enforce the lowest degradation of a human being, that of being made the instrument of an animal function contrary to her inclinations.[81]

Mill's thesis was bitterly contested. In 1873, for example, Fitzjames Stephen addressed *The Subjection of Women* with the explanation that it was a "work from which I dissent from the first sentence to the last".[82] In turn, Millicent Garrett Fawcett described Stephen's *Liberty, Equality, Fraternity* as

> the latest revelation of the Gospel according to St. Stephen. Upon nearly all the most important subjects of contemporary politics, upon the gravest questions of religion and metaphysics, we are fully instructed what we ought to do, and how we ought to think.[83]

Stephen's work illustrates his personal understanding of the role and purpose of the English criminal law, as well as his ideological and political engagement against the discourses of modern democracy.

Even prior to his conversion to the position of absolutely irrevocable consent in marriage, Stephen J saw Mill's doctrine of sex equality as resting on "an unsound view of history, an unsound view of morals, and a grotesquely distorted view of facts, and I believe that its practical application would be as injurious as its theory false".[84] For Stephen J, the argument for equality was misguided and dangerous in general. He believed that unequal treatment of the sexes in particular is natural; that it serves the important purpose of maintaining social stability, and moreover that the "actually existing generation of women do not dislike their position".[85] In fact, he seems to have considered the subject a distasteful choice for public discussion in general, criticising Mill on the grounds that

> there is something – I hardly know what to call it; indecent is too strong a word, but I may say unpleasant in the direction of indecorum – in prolonged and minute discussions about the relations between men and women, and the characteristics of women as such.[86]

Stephen's antipathy to Mill was reciprocated. John Stuart Mill was critical of Stephen's outlook and character in general. In 1869 Mill wrote to Thomas Cliffe Leslie that

[80] *Ibid*, 504.

[81] *Ibid.*

[82] Sir James Fitzjames Stephen, *Liberty, Equality, Fraternity* (1873), edited by RJ White, (Cambridge University Press, Cambridge, 1967), 188.

[83] Millicent Garrett Fawcett, *Mr Fitzjames Stephen on the Position of Women*, (London, Macmillan & Co, 1873) 5.

[84] Sir James Fitzjames Stephen (1873), *op cit*, 191.

[85] *Ibid*, 189.

[86] *Ibid*, 190.

> I agree with you in exceedingly disliking the insolent and domineering affectations of Fitzjames Stephen. In political economy he is exceedingly ignorant, but not less presumptuous. On other matters, however he is able to do some useful work, and he is undoubtedly a clever man. My daughter begins to have some doubts whether he is thoroughly an honest man, either in word or deed. It is certain that he says and does oddly inconsistent things. He is always brutal, even at his best; that however apparently in him a radical defect of temperament, which if he is otherwise an honest worker for good, might have to be overlooked.[87]

And Mill went on to satirise the vanity of Stephen:

> It is certain that he is very vain, and that may be the cause of his many defects in which vanity is not apparent, as of the boyish boast that he always goes to sleep at the P.E. Club. Has he then so much time to spare; or does he mean that he comes because he cannot get sleep elsewhere? But he is full of this sort of fanfaronade, which is offensive enough, but which we may excuse if he is in earnest about anything. One cannot help but hoping he is because he is clever enough to do a good deal of good or mischief.[88]

The "brutal" views of Stephen J on equality and the criminal law and the sanctity of marriage help to shed light on his judgment in *Clarence.* With this insight, his motivation and intention in *Clarence* can be seen more clearly. Stephen J was opposed to the doctrine of equality, and sexual equality in particular. He believed that women were inferior to men in strength of body and mind, and that marriage was built solidly on this fact. Importantly, he did not consider equality to be particularly relevant to justice, and he did not consider it the role of the criminal law to regulate "personal" relations. He wrote in response to John Stuart Mill:

> I think that equality has no special connection with justice, except in the narrow sense of judicial impartiality; that it cannot be affirmed to be expedient in the most important relations of social life; and that history does not warrant the assertion that for a great length of time there has been a continual progress in the direction of the distinctions between man and man, though it does warrant assertion that the form in which men's natural inequalities display themselves and produce their results changes from one generation to another, and tends to operate rather through contracts, made by individuals that through laws made by public authority for the purpose of fixing the relations between human beings.[89]

In the decades after the treatises of Mill and Stephen, women's rights were addressed by Parliament as well as by the courts. In particular, the 1880s saw the views on equality of Stephen J routinely challenged. In 1880 William Gladstone was elected on a platform of equality. In 1882 the Married Women's Property Act was passed, under which married women possessed the same rights over their property as unmarried women, thereby allowing a married woman to retain ownership of property. In 1884 the Matrimonial Causes Act allowed a wife deserted by an adulterer to petition for divorce immediately, rather than waiting two years. This Act might have had repercussions for women beyond those who used the legislation to escape unwanted marriages. Judith Knelman identifies the reporting of divorce in the press in the 1880s as testing "middle-class morality" more generally in its indication that "not all wives

[87] John Stuart Mill, "Letter to T E Cliffe Leslie, 8 May 1869", *The Later Letters of John Stuart Mill 1849–1873*, collected works, edited by Frances E Minkea and Dwight N Lindley, (University of Toronto Press, Toronto, 1963) 1600.

[88] *Ibid.*

[89] Sir James Fitzjames Stephen (1873), *op cit*, 210.

were happy or even faithful".[90] In 1885 the Criminal Law (Amendment) Act raised the age of female heterosexual consent from 12 to 16, a move viewed variously as benevolently paternalistic, repressive and misguided,[91] or as primarily in *men's* cynical interest.[92] The Act also contained a clause that recognised as rape the obtainment of sex by impersonating a husband.[93] In 1886 as a result of large scale lobbying, the Contagious Diseases Acts of 1864 were repealed, ending the degrading and intrusive ritualised detainment and inspection of so-called "common prostitutes". Against the grain of these moves, Stephen J changed his views on Hale's irrevocable consent doctrine, and the damning decision of *Clarence* was delivered in 1888.

## THE ROLE OF PRECEDENT

Critical historical analysis is imperative in assessing the role of precedent in criminal law, yet it is often overlooked. In particular, individual personalities of the bench and their personal contributions to interpreting or making law deserve attention. The "old" view of the judges' role was that they were merely "declaring" the existing law. In 1892 Lord Esher voiced the idea that "there is, in fact, no such thing as judge-made law, for the judges do not make the law, though they frequently have to apply existing law to circumstances as to which it has not previously been authoritatively laid down that such law is applicable".[94] Yet judges do make law, influenced by their own prejudices, experiences and moments in history. SA Farrer notes the profound limitations of the criminal law tradition that does not accommodate an analysis of the role of history. For Farrer,

> [i]n discussions on English criminal law, evidence and procedure, history appears to have only marginal relevance. English lawyers tend to work within a practical tradition that seeks to unravel and solve contemporary problems rather than to dwell on the past. So within this frame, when history is discussed it is often presented as posited, objective, uncomplicated and uncontested terrain. Yet history does not consist of objective accounts in which the facts speak for themselves. Rather, facts are selected, interpreted and constructed within evaluative frameworks that generate an image of history.[95]

David Ormerod and Michael Gunn note in reference to *Clarence*, "longevity of a case is sometimes a reason in itself for refusing to overrule it".[96] But they also note that the judgment reflects "somewhat outmoded attitudes to sexuality, and therefore the case deserves to be re-examined".[97] Noting Farrer's emphasis on the role of history in informing judgments that are neither "posited, objective, uncomplicated or uncontested", the question arises concerning *Clarence*: if the reasoning is based on prejudice and an outdated ideological stance, is any part of the judgment, its content or logic,

[90] Judith Knelman, "Why Can't a Woman be More Like a Man? Attitudes to Husband–Murder 1889–1989", in Rowbotham and Stevenson (eds), *Behaving Badly; Social Panics and Moral Outrage – Victorian and Modern Parallels*, (Aldershot, Ashgate, 2003) 197.

[91] Nicki Roberts, *Whores in History: Prostitution in Western Society*, (Harper Collins, London, 1992) 257.

[92] Kim Stevenson, "Observations on the Law Relating to Sexual Offences: The Historic Scandal of Women's Silence", (1999) 4 *Web Journal Of Current Legal Issues*.

[93] Notoriously it also included the Labouchere Amendment that criminalised all gay male sex acts other than anal sex (which was already included in the crime of buggery).

[94] *Wills v Baddeley* [1892] 2 QB 324, *per* Lord Esher, 326.

[95] SA Farrer, "Myths and Legends: An Examination of the Historical Role of the Accused on Traditional Legal Scholarship: A Look at the 19th Century", (2001) 21 *Oxford Journal of Legal Studies*, 331.

[96] David Ormerod and Michael Gunn, "Criminal Liability for the Transmission of HIV", (1996) 1 *Web Journal of Current Legal Issues*, 13.

[97] *Ibid.*

retrievable? Is there any wisdom to be gleaned from a case that was recently described by the Canadian Supreme Court in *Cuerrier* as "harsh and antiquated",[98] or is *Clarence* beyond redemption? I conclude the latter. As Gonthier and McLachlin JJ argue in *Cuerrier*, "it is the proper role of the courts to update the common law from time to time to bring it into harmony with the changing needs of society. This applies to the common law concept of fraud in relation to assault".[99] This observation applies to *Clarence*.

Although Philpot J in *Dica* argued that *Clarence* has been "thoroughly undermined", the judgment might yet rise from the dead, as it has done miraculously on a number of occasions since it appeared to have been "thoroughly undermined" by the House of Lords in 1992. An appeal in *Dica* is pending and, as I have noted, although it is no longer an authority on sex and disease, *Clarence* has not been explicitly dismissed by the courts as an authority on consent and fraud in other areas. I argue that the *ratio* of *Clarence* cannot, in fact, be extracted from the prejudice that informed its arguments about marital rape. The majority reasoning exhibits fundamental assumptions about the nature of that harm which cannot be identified easily as physical violence.

*Clarence* is a callous judgment. Stephen J himself noted, in writing of the history of English law, that the English common law tradition does not recognise the gravity of personal harm: "the extraordinary leniency of the English criminal law towards the most atrocious acts of personal violence forms a remarkable contrast to its extraordinary severity with regard to offences against property".[100] I would add that, in particular, personal harm that is not predicated on grave "physical violence" has been inadequately conceptualised and addressed in the law. *Clarence* is a callous judgment that gives no weight to personal autonomy in its assessment of consent and harm. The simple question of consent concerns autonomy, and autonomy is not the focus of *Clarence*.

## CONCLUSIONS – LEARNING FROM CLARENCE

There are two important points to be made when considering *Clarence* and its legacy in terms of assessing how we understand the role of precedent in the common law. The first is that early legal opinion was not sacrosanct among the late Victorian bench. As the judgment in *Jackson* shows, the modern law was evolving and the judges of the day were capable of reflection on and critical appraisal of early authority, the doctrines from which they seemed to exercise their authority to choose. The choice of Hale's opinion was not a given. It might easily have been rejected in this climate, as *Jackson* suggests. The second point concerns the position of women and the emergent discourse of sex equality. As noted, the legislature had responded to demands for repeal and implementation of various Acts concerning women's rights, particularly in the 1880s. The legislative responses followed public debates not only over women's sexuality, but also concerning the nature and desirability of equality.[101]

The political implications of choosing to confirm Hale's opinion should not have been lost on the Queen's Bench that presided over *Clarence*, particularly Fitzjames Stephen, who was explicit about his newly developed and unconditional support for

[98] *R v Cuerrier* [1998] 2 SCR 371, *per* Corey J.

[99] *Ibid, per* Gonther and McLachlin JJ.

[100] Sir James Fitzjames Stephen, *A History of the Criminal Law of England*, (Macmillan & Co, London, 1883), 109.

[101] For example, Married Women's Property Act 1882, Matrimonial Causes Act 1884, Criminal Law (Amendment) Act 1885 and in 1886 the Contagious Diseases Acts of 1864 were repealed.

Hale's opinion. The judgment cannot be understood in isolation from the contemporary political climate, nor merely as a product of "Victorian values", for such values were not homogeneous and were contested strongly even within the ruling class, as the debates between Mill and Stephen show. Introducing Hale's opinion in the *Clarence* trial involved a conscious and tactical choice by the defence, and affirming Hale's opinion entailed a political choice by the members of the bench. As well, the general denial of criminal sexual harm by the majority in *Clarence,* aside from the marital rape exemption, was a well thought-out and purposeful response.

In this light, the abstract or dislocated use of common law precedent becomes questionable. Isolating the agenda and the surrounding environment of the *Clarence* decision provides important insight into the motivation and the intention of the members of the bench in making their judgments. The common law relies on precedent, often uncritically. It needs to be established whether it is valid to abstract a common law decision from its origins and continue to apply it as relevant beyond contemporary values. Critical analysis of *Clarence* questions the worth of transporting dislocated and abstract rulings from their political environments into the present day as persuasive authority, particularly in the dynamic, contested and developing area of sexual rights.

In criticising *Clarence* I would not argue for the prosecution of HIV transmission in the manner of *Dica.* Rather I argue that the dynamic, contested and evolving area of sex and the criminal law warrants a more appropriate authority than *Clarence* or its legacy. In this article I have highlighted what I identify as the motivation of the bench in *Clarence*: a fundamental lack of interest in the harm of sex, particularly for women, particularly within marriage. I have illustrated the personal agenda of James Fitzjames Stephen with regard to women, equality and the criminal law, to conclude that the judgment is so bound up with this agenda that its logic cannot be salvaged for use in discerning harm in sex, or to informing understandings of consent in general. HIV has brought *Clarence* once more into the spotlight. Now, more than ever, fraud in sex is potentially deadly and needs to be reviewed and addressed comprehensively. It is time to put *Clarence* to bed.

# REFORMING INSOLVENCY LAW: A COMPARATIVE STUDY OF SCOTLAND AND SOUTH AFRICA

*DONNA MCKENZIE SKENE**

## INTRODUCTION

Insolvency law in the UK has recently been subject to considerable scrutiny and reform as part of the present government's policy for increasing productivity and enterprise. The Enterprise Act 2002 ("EA 2002") introduced major reforms to company insolvency law in both Scotland and England and Wales which came into force on 15 September 2003 and major reforms to non-company insolvency (bankruptcy) law in England and Wales which came into force on 1 April 2004; Part I of the Debt Arrangement and Attachment (Scotland) Act 2002 ("DAA(S)A 2002") introduced in Scotland a debt arrangement scheme ("DAS") designed to allow individual debtors with multiple debts to repay them in a managed way as an alternative to the formal insolvency process of sequestration which came into force on 30 November 2004; and the Scottish Executive is currently consulting on reform of bankruptcy law in Scotland.[1]

In South Africa, the South African Law Commission ("SALC") published its *Report on Review of the Law of Insolvency* in 2000.[2] This report ("the SALC report") set out recommendations for what it described as mainly technical reforms to non-company insolvency law in South Africa and also contained proposals designed to provide a workable alternative to the formal insolvency process of sequestration for individual debtors with multiple debts. In the same year, the Centre for Advanced Corporate and Insolvency Law at the University of Pretoria ("CACIL"), acting on a remit from the Standing Advisory Committee of Company Law, produced proposals for a new uniform insolvency law based on the SALC's proposals for reform of non-company insolvency law but incorporating reformed company insolvency law provisions including business rescue provisions. These proposals have been accepted in principle. In 2002, CACIL produced a further (interim) report on the administration order procedure available to individual debtors as an alternative to sequestration and reform of the administration order procedure is now being considered by the SALC, which is expected to issue a Discussion Paper in the spring of 2005. In August 2004, the South African Department of Trade and Industry published for public comment a Consumer Credit Bill which includes provisions for debt re-structuring for consumer debtors and it is anticipated that legislation based on this will be introduced in the first half of 2005.

This article considers the recent and proposed reforms in these two jurisdictions.[3] Such a comparison is particularly appropriate for a number of reasons. Firstly, both

* Senior Lecturer, School of Law, University of Aberdeen. The author gratefully acknowledges the support of the British Academy in funding a research visit to the University of Pretoria for the purposes of researching the South African material for this article and the Faculty of Law at the University of Pretoria for providing excellent research facilities and other invaluable assistance during her stay. Among many others, particular thanks go to Professor Andre Boraine, Director of the Centre for Advanced Corporate and Insolvency Law at the University of Pretoria; Professor David A Burdette, University of Pretoria; Melanie Roestoff, University of Pretoria; Professor Alastair Smith, University of South Africa; Roger Evans, University of South Africa and Tienie Cronje, South African Law Commission, for taking so much time to make her welcome and assist her in her research; further thanks are due to Professor Boraine and Professor Burdette for reading an earlier draft of this article – the usual disclaimer applies.

[1] Company insolvency law is substantially the same in the two jurisdictions and is in most respects reserved to the UK parliament, while bankruptcy law in Scotland remains distinct from bankruptcy law in England and Wales and is in most respects devolved to the Scottish Parliament.

[2] South African Law Commission, Project 63, *Review of the Law of Insolvency* (2000).

[3] It should be noted that the law relating to international insolvencies has also been subject to reform in both jurisdictions, but considerations of space preclude discussion of these reforms and this article is accordingly confined to the reforms relating to purely domestic insolvencies in the two jurisdictions.

jurisdictions are classed as mixed legal systems. Secondly, the insolvency laws of both jurisdictions are currently structured along similar lines in that there are separate procedures for company debtors[4] and non-company debtors[5] respectively. Thirdly, South African company insolvency law was based originally on company insolvency law in England and Wales and retains many recognisable features of the latter while company insolvency law in Scotland, too, was and remains in most respects the same as that in England and Wales. Fourthly, the main aims of reform in both jurisdictions are broadly similar, including the provision of effective business rescue procedures and alternatives to formal insolvency processes for individual debtors.

The article commences with a brief description of the law in both jurisdictions before moving on to consideration of the reforms/proposed reforms.

## SCOTS LAW

### *General Structure Of Insolvency Law In Scotland*

Scots law provides different formal insolvency procedures for company and non-company debtors, although the provisions of company insolvency law and bankruptcy law have been harmonised where possible.[6] The only formal insolvency procedure applicable to non-company debtors, that is individuals (whether in business or consumers), partnerships (including dissolved and limited partnerships but not limited liability partnerships ("LLPs"), which are treated for insolvency purposes in the same way as company debtors), trusts and other corporate and unincorporated bodies is sequestration under the Bankruptcy (Scotland) Act 1985 ("B(S)A 1985").[7] There are a number of formal insolvency procedures applicable to companies and, in suitably modified form, to LLPs, under the Insolvency Act 1986 ("IA 1986") and the Companies Act 1985 ("CA 1985"). These are: company voluntary arrangements under Part I of IA 1986 ("CVAs"); compromises or arrangements under section 425 of CA 1985 ("section 425 arrangements"); receivership; administration; and liquidation (also referred to as winding up). There are also a number of alternatives to formal insolvency procedures.

### *The Administration Of Insolvency Procedures*

The Accountant in Bankruptcy ("AIB"), now established as a government agency, may undertake the administration of sequestrations either directly or through appointed agents[8] and also, *inter alia*, exercises a supervisory role in relation to sequestrations generally.[9] Sequestrations other than those administered by the AIB and all company insolvency proceedings are administered by licensed insolvency practitioners ("IPs") or, in the case of CVAs, persons specifically authorised to carry out CVA work. IPs are generally accountants or solicitors and in order to become licensed to act as IPs are

[4] Including, in the UK, limited liability partnerships and in South Africa, close corporations: see further below.

[5] These include, in both jurisdictions, individuals, partnerships (with the exception of limited liability partnerships in the UK), trusts and other unincorporated bodies: see further below.

[6] See Scottish Law Commission *Report on Bankruptcy and Related Aspects of Insolvency and Liquidation*, Scot Law Com No 68 (1982) ("the Scottish Law Commission Report"), para 1·6 and the *Report of the Review Committee on Insolvency Law and Practice*, Cmnd 8558 (1982) ("the Cork Report"), para 1980. One example of such harmonisation is the provisions for challenge of certain types of transaction carried out prior to the commencement of insolvency proceedings.

[7] There is no Scottish equivalent to individual voluntary arrangements under Part VIII of the Insolvency Act 1986 ("IVAs") which are available only in England and Wales.

[8] The AIB's Annual Report for 2003/2004 discloses that the AIB was appointed as interim trustee in 89% of sequestrations in that year and that around 75% of these cases were allocated to agents; figures for the previous year were similar.

[9] The AIB has only had a role in the actual administration of estates since 1993: see further below.

required to fulfil a complex set of requirements including educational and experiential requirements.[10] Persons specifically authorised to carry out CVA work must also fulfil a number of requirements including educational and experiential requirements.[11]

*Alternatives To Formal Insolvency Procedures*

Most debtors will try to avoid formal insolvency procedures where possible and may try to negotiate informal arrangements with individual creditors for that purpose e.g. extension of time for payment, instalment payments. Other possibilities are:

(i) A composition at common law.[12] Any debtor may enter a composition with some or all creditors in terms of which the participating creditors agree to accept less than full payment of their debts (often in instalments) and the debtor receives a discharge. Creditors must be treated rateably. These are contractual arrangements to which creditors must consent and therefore tend to be of limited application in practice.

(ii) The DAS. The DAS came into force on 30 November 2004 and is designed to allow individual debtors with multiple debts to enter a debt payment programme ("DPP") for repayment of their debts while protected from enforcement action. An application for approval of a DPP is initiated by the debtor after he or she has received appropriate money advice. A DPP will provide for the debtor to make a single periodic payment, which may be deducted at source from the debtor's earnings and which will be based on the debtor's surplus income after payment of essential outgoings including ongoing payments to creditors, to an approved payments distributor for distribution among the creditors included in the DPP in accordance with its terms. The debts which may be included in a DPP are defined widely and include arrears of secured debt. Creditors must consent to a DPP, but there is provision for binding dissenting creditors. There is no provision for discharge of the debtor on less than full payment, although individual creditors may, and are encouraged to, agree to waive interest on or compound their debts. Most enforcement action other than the enforcement of a security is stopped during the currency of a DPP, while the debtor is restricted from taking on new credit. A DPP may be varied on the application of the debtor or a creditor and may be revoked in defined circumstances. There is a public register of DPPs. It is too early to tell how many debtors will utilise the DAS in practice.

(iii) Trust deeds for creditors. A trust deed for creditors is a voluntary deed by a debtor conveying specified assets to a named trustee to be administered for the benefit of creditors and the payment of debts. The trust deed may also provide for contributions from income during the currency of the trust deed. Creditors who do not accede to a trust deed, however, are not bound by it unless it becomes protected under the B(S)A 1985. To become protected, the trust deed must convey to the trustee the same assets as would vest in a trustee in sequestration; creditors are given an opportunity to object to the trust deed; if a specified minority do not object within specified time limits, the trust deed becomes protected and dissenting creditors are then bound and there are limited rights of challenge to the deed. Trust deeds may, in theory, be granted by companies, but it is thought that a trust deed granted by a company may not become protected and would therefore be of limited application in practice.[13] In

[10] See Part XIII of IA 1986 and associated delegated legislation, in particular the Insolvency Practitioners Regulations 2005, SI 2005/524, as amended. The delegated legislation has just been reviewed, but with a view only to "fine-tuning" and modernising what is regarded as a fundamentally successful regime rather than its wholesale change: see consultation package available at http://www.insolvency.gov.uk.

[11] See Part XIII of IA 1986 as amended by the Insolvency Act 2000.

[12] These should not be confused with compositions under s 56 and Sch 4 of B(S)A 1985, which provide for discharge of the debtor on composition following sequestration: such compositions are generally referred to as judicial compositions and are discussed further below.

[13] Trust deeds which do not become protected are rare.

practice, trust deeds operate in much the same way as sequestration but without the formalities and some of the more serious consequences of sequestration and have thus become an increasingly popular alternative to sequestration.[14]

*The Formal Insolvency Procedures Available In Scots Law*

Sequestration is a judicial procedure whereby the debtor's assets are placed in the hands of a trustee for realisation and distribution among the creditors. It is initiated by a petition to the appropriate court which may be made, *inter alia*, by the debtor or a qualified creditor or creditors. Where it awards sequestration, the court also appoints an interim trustee who administers the sequestration process until a permanent trustee ("the trustee") is elected by the creditors or, absent such election, appointed by the court. Sequestration vests the assets of the debtor, with limited exceptions, in the trustee, who sells or otherwise realises them and then distributes them amongst the creditors according to their respective rights. There is provision for the election of up to five commissioners (creditors elected by the creditors themselves) who play an important role in advising the trustee and sanctioning certain of his actions; where no commissioners are elected, these functions are carried out by the AIB and/or the court. In certain defined cases, principally those where there are few or no assets, modified procedures apply with a view to saving time and costs.[15] Unless it is deferred by the court on cause shown following an application by the trustee or a creditor, the debtor is entitled to an automatic discharge three years after the commencement of sequestration which relieves the debtor of liability for pre-sequestration debts with limited exceptions. There is also provision for discharge on composition by the debtor at any time after the commencement of sequestration. The administration of the sequestration itself may continue beyond the debtor's discharge, in which event the debtor has a continuing obligation to co-operate in the sequestration process; on its completion, the trustee also obtains a discharge and the sequestration process is effectively brought to an end.

A CVA is a binding "composition in satisfaction of its debts or ... scheme of arrangement of its affairs" between the company and its creditors. Introduced by the IA 1986, it may be initiated by the directors of the company or, where the company is already in administration or liquidation, by the administrator or liquidator. Under new provisions introduced by the Insolvency Act 2000 ("IA 2000"),[16] where the directors of an eligible company wish to initiate the CVA procedure, they may obtain a short moratorium, monitored by an IP, protecting the company from enforcement action by creditors to enable them to take the necessary steps to do so.[17] The procedure for concluding a CVA varies slightly depending on who is initiating the procedure and whether or not a moratorium has been obtained, but the essential steps are the same in all cases. A proposal is drawn up setting out the details of the proposed arrangement and nominating an IP as nominee to oversee the procedure. Meetings of the company and its creditors are then called to consider the proposal. The respective meetings may approve the proposal (with or without modifications) by a defined majority in each

[14] The AIB's Annual Report for 2003/2004 gives the figures for sequestrations and protected trust deeds for the years 2000–2004 as follows: in 2000 there were 3,185 sequestrations and 2,353 protected trust deeds; in 2001, 2,938 sequestrations and 2,943 protected trust deeds; in 2002 there were 3,193 sequestrations and 4,011 protected trust deeds; in 2003 there were 3,228 sequestrations and 5,363 protected trust deeds and in 2004 there were 3,309 sequestrations and 5,669 protected trust deeds.

[15] There are currently two types of modified procedure: Schedule 2 procedure and summary administration.

[16] The provisions of that Act relating to CVAs came into force on 1 January 2003.

[17] IA 1986, s 1A and Sch A1 as added by IA 2000. An eligible company is one falling within the definition of a small company as set out in CA 1985, s 247(3) and not otherwise excluded by IA 1986, Sch A1. The effect of the moratorium is to stop most enforcement action by creditors, including secured creditors.

case;[18] where the decisions of the two meetings differ, however, the creditors' decision takes precedence.[19] If approved, the CVA takes effect immediately and binds all those who were entitled to vote at the meeting or would have been so entitled if they had received notice of it, subject to a limited provision for challenge. The implementation of the CVA is overseen by the nominee or his or her replacement, now known as the supervisor. The CVA will typically provide for a creditors' committee, but it is not compulsory. There is, by design, little court involvement in this procedure although applications may be made to the court in certain circumstances.

A section 425 arrangement is a compromise or arrangement between a company and its members and/or creditors and/or any class of its members and/or creditors. Originally designed for insolvency situations, the procedure is no longer so restricted, but may still be utilised by an insolvent company including a company in receivership, administration or liquidation. It is initiated by a petition to the appropriate court requesting permission to call meetings of the members, creditors and/or any class(es) of members and/or creditors with whom the company wishes to enter into an arrangement. Where the requisite permission is granted, the relevant meetings are called and may accept the proposed arrangement by the requisite majority.[20] The matter is then referred back to the court, whose sanction is also required; where this is granted, the arrangement takes effect and binds all members/creditors/classes of members and/or creditors with whom it was concluded and is implemented accordingly. Similar in a number of respects to CVAs, in practice a CVA would generally be chosen in preference to a section 425 arrangement because of the minimal amount of court involvement in the former as compared to the latter.[21]

Receivership is the primary mechanism available to the holder of the particular type of security known as a floating charge for enforcement of that security, and although included here as a formal insolvency procedure, it differs in important respects from the other formal insolvency procedures. Where a company that has granted a floating charge is in default, the floating chargeholder may appoint, or apply to the court for the appointment of, a receiver, whereupon the floating charge crystallises and becomes a fixed security over the property affected by it. Where the floating charge affects all or substantially all of the company's property, the receiver is known as an administrative receiver and must be an IP. The receiver takes control of the property affected by the floating charge and has power to manage and realise it with a view to paying the floating chargeholder's debt. There is provision for a committee of creditors, but it has a very limited role. The receiver pays any monies realised to the floating chargeholder in satisfaction of his or her debt after first paying certain other specified creditors, but is not obliged to deal with the claims of other creditors. On the basis that receivers owe only limited duties to creditors other than the floating chargeholder and that receivership is not a collective procedure because receivers are required to deal only with the claims of certain specified creditors, the EA 2002 introduced new provisions to prevent the appointment of an administrative receiver under a floating

[18] In the case of the creditors' meeting, by a majority of at least three quarters in value of the creditors present or represented and voting, in person or by proxy and in the case of the company meeting, by a simple majority in value of those voting in person or by proxy. For this purpose, the value of a secured creditor's claim is the value of any unsecured balance of his claim only. The meetings may not, however, approve a proposal affecting certain rights of secured or preferential creditors without their specific consent.

[19] IA 1986, s 4A as added by the IA 2000. In such a case, there is provision for a member of the company to apply to the court, which may order the decision of the company meeting to have effect instead of the decision of the creditors' meeting or make such other order as it sees fit.

[20] A majority in number representing three-fourths in value of those present and voting either in person or by proxy.

[21] A s 425 arrangement may be regarded as more appropriate in particular circumstances, however, and historically, it had the advantage over a CVA of binding unknown creditors, a problem with CVAs finally resolved by the Insolvency Act 2000.

charge created after 15 September 2003 except in certain limited circumstances; it remains possible, however, to appoint an administrative receiver under any floating charge created prior to that date.

Administration is the primary rescue procedure available to companies in financial distress. Introduced by the IA 1986, the legislation originally provided for administration to be entered for one or more of four separate purposes: the survival of the company and the whole or part of its undertaking as a going concern; the sanctioning of a section 425 arrangement; the approval of a CVA; and a more advantageous realisation of assets than would be available on winding up. All administrations were initiated by an application to the appropriate court and an interim moratorium took effect on presentation of the application. The interim moratorium did not, however, prevent a floating chargeholder from appointing a receiver, and where an administrative receiver had been appointed (whether before or after the application), the court could not appoint an administrator unless the floating chargeholder consented or the court was satisfied that the floating charge would be invalid or subject to challenge. Thus, a floating chargeholder could effectively veto an administration. On the making of an administration order, an administrator would be appointed to manage the affairs, business and property of the company and a full moratorium would come into effect. The administrator would then take over the running of the company and formulate proposals for achieving the purpose(s) for which the administration order had been granted. These proposals would be voted on by the creditors: if approved by the requisite majority,[22] the proposals would be implemented by the administrator, otherwise the administration would generally be brought to an end. A creditors' committee could be established to assist the administrator in discharging his functions. As a result of concerns that the administration procedure was being under-utilised and an insufficient number of companies rescued, coupled with a desire to move away from administrative receivership in favour of a collective procedure, the EA 2002 made a number of significant changes to the administration procedure which apply to administrations commenced after 15 September 2003. Thus, it is now possible for an out-of-court appointment of an administrator to be made by a floating chargeholder *or* by the company or its directors and an interim moratorium comes into effect as soon as that appointment process is commenced. An administrator may still be appointed on application to the court, and in these cases the floating chargeholder's veto has been retained, but in time this will become less significant as a result of the new restrictions on the appointment of administrative receivers discussed above. The original four purposes of administration have been replaced with one composite purpose with a hierarchy of objectives, the primary objective being the rescue of the company as a going concern. A time limit of one year on administration has been introduced, but this may be extended by the creditors or the court. Also introduced for the first time are specific powers for administrators to make distributions to creditors, including secured creditors, within administration itself.

Liquidation (or winding up) is the company equivalent of sequestration. The company's assets are placed in the hands of a liquidator to be realised and distributed to the creditors and, in a solvent liquidation, the members. The company is then dissolved. There are two types of liquidation: voluntary liquidation and compulsory liquidation (also referred to as winding up by the court). Voluntary liquidation is initiated by resolution of the members and may have no or little court involvement. There are two types of voluntary liquidation: members' voluntary liquidation, where

[22] In this case, a simple majority in value of those voting in person or by proxy. The value of a secured creditor's claim is the value of any unsecured balance of his claim only.

the directors have made a statutory declaration of solvency to the effect that the company will be able to pay all its debts with appropriate interest within a period not exceeding 12 months; and creditors' voluntary liquidation, where no such statutory declaration has been made. A members' voluntary liquidation is, therefore, normally a solvent liquidation although there is provision for it to be converted into a creditors' voluntary liquidation where the liquidator forms the view that the company will not in fact be able to pay its debts and interest within the period stated in the statutory declaration. Compulsory liquidation is initiated by application to the appropriate court. On the making of the winding up order, an interim liquidator is appointed who administers the liquidation pending the election or appointment of a liquidator. In all types of liquidation, the liquidator gathers in the company's assets and sells or otherwise realises them prior to distributing them: in a solvent liquidation, the liquidator pays the creditors and then distributes the remaining assets to the members according to their rights; in an insolvent liquidation, the assets are distributed to the creditors in accordance with their respective rights. There is provision for the election of a liquidation committee in creditors' voluntary and compulsory liquidations that plays an important role in the liquidation process akin to that of the commissioners in a sequestration. On completion of the liquidation process, the company is dissolved.

In all types of insolvency procedure, funds are distributed to creditors according to their respective rights. It is a recognised principle of insolvency law in Scotland that creditors should be treated *pari passu*, but this principle is imperfectly applied and certain categories of creditors are given priority over others. In particular, certain debts due to the Crown for unpaid taxes etc and certain debts due to employees have traditionally been given preferential status in insolvency.[23] The EA 2002, however, abolished Crown preference in all types of insolvency procedure and, in order to ensure that the benefit of this change was directed to unsecured creditors, introduced a provision requiring receivers, administrators and liquidators to set aside a proportion of the funds destined for floating chargeholders to be distributed to the unsecured creditors in the receivership, administration or liquidation, subject to limited exceptions. This provision, however, affects only floating charges created after 15 September 2003.

## SOUTH AFRICAN LAW

### *General Structure Of Insolvency Law In South Africa*

Like Scotland, South African law provides different formal insolvency procedures for company and non-company debtors, although many of the provisions of company and non-company insolvency law are the same or similar.[24] The formal insolvency procedure applicable to non-company debtors, that is individuals (whether consumers or in business), partnerships (with the exception of partnerships the members of which are all juristic persons) and entities, associations of persons, companies or bodies corporate which are not capable of being wound up under the Companies Act 61 of 1973 ("Act 61 of 1973") is sequestration under the Insolvency Act 24 of 1936 ("Act 24 of 1936"). There are a number of different procedures available to companies and close corporations. A company, but not a close corporation, may enter into a compromise or arrangement under section 311 of Act 61 of 1973 ("a section 311 compromise") or

[23] Such debts are referred to as preferred debts in the context of sequestration and preferential debts in the context of company insolvency proceedings, but are essentially the same in each case.

[24] Partly because many of the provisions of non-company insolvency law are applied directly to company insolvency law.

be made subject to judicial management under Act 61 of 1973. Both companies and close corporations may be wound up under the provisions of the Act 61 of 1973 and the Close Corporations Act 69 of 1984 ("Act 69 of 1984") respectively, the latter Act effectively applying many of the provisions of the former to the winding up of close corporations. There are also a number of alternatives to formal insolvency procedures.

*The Administration Of Insolvency Procedures*

The Master appointed to each of the areas of the provincial divisions of the High Court under the Administration of Estates Act 66 of 1965 has an important role in the administration of all types of insolvency proceedings in South Africa as well as a general supervisory role similar in many respects to that of the AIB in non-company insolvencies in Scotland, but the main part of the administration in any insolvency is undertaken by the individual trustees, liquidators or judicial managers appointed in the proceedings. Currently, there is no equivalent of the UK regulatory regime requiring persons acting in these capacities to be licensed or otherwise authorised to carry out such work,[25] although one of the proposed reforms currently under consideration is the introduction of a requirement that such persons should at least be members of an authorised body.[26]

*Alternatives To Formal Insolvency Procedures*

There are a number of alternatives to formal insolvency proceedings.

(i) Informal arrangements with individual creditors.

(ii) Common-law compromise. Similar to a composition at common law in Scotland, this involves a debtor entering into a written agreement with creditors to pay certain dividends on the creditors' claims on condition of a release from remaining debts. It may be entered into by a debtor even after a provisional sequestration or winding up order has been made,[27] in which case the agreement will include the provisional trustee/liquidator and provide for the discharge of the relevant order. As in Scotland, however, it is contractual in nature and therefore of limited practical use due to the need to obtain the consent of all creditors.

(iii) Administration orders under the Magistrates' Courts Act 32 of 1944 ("Act 32 of 1944"). Similar to county court administration orders under the County Courts Act 1984 in England and Wales, there is no Scottish equivalent, although the new DAS contains a number of similar features. Available to natural persons only, administration has been described as a modified form of insolvency suited to dealing with relatively small estates where the cost of sequestration would exhaust the estate and as having the aim of assisting a debtor over a period of financial embarrassment without the need for sequestration.[28] An administration order may be made following a judgement against the debtor if he cannot immediately pay the debt or on an application by the debtor on the basis that he cannot pay his debts generally,[29] but is only available where the total debts do not exceed R50,000.[30] The application is made on a prescribed form and following notice to creditors, a hearing takes place at which creditors are given an opportunity to object to the making of the order. If made, the

[25] Although there are a number of provisions disqualifying certain persons from acting in such capacities; similar disqualifications are part of the UK regime.

[26] This is discussed further below.

[27] See further below.

[28] LAWSA, Vol 3(2), para 233 and references there cited.

[29] Act 32 of 1944, s 74.

[30] "Debts" in this context means debts due and payable, which excludes obligations to pay money *in futuro*: *Cape Town Municipality v Dunne*, 1963 1 SA 741 (C).

order will specify a weekly or monthly amount representing approximately the difference between the debtor's expected income and certain prescribed "necessary expenses" to be paid by the debtor to the administrator appointed by the court for distribution to the creditors and the court may order an emoluments or garnishee attachment order to facilitate the payments. The administrator may also realise assets of the debtor with the permission of the court. With some exceptions, individual debt collection measures by creditors whose debts are included in the administration are suspended during the currency of the administration order, but *in futuro* debts, which includes ongoing payments under, for example, hire purchase contracts, are not capable of being included in an administration order, which effectively limits its utility for many debtors; furthermore, administration is not a bar to sequestration itself. There is no restriction on a debtor under administration incurring further debt, although it is an offence not to disclose the existence of the administration order when doing so. The administration order may be suspended, amended or rescinded in defined circumstances, including a change of the debtor's circumstances. Debts must be paid in full under an administration order: there is no provision for discharge of the debtor on less than full payment.

*The Nature Of The Formal Insolvency Procedures Available In South African Law*

Sequestration is a judicial procedure for placing a debtor's assets in the hands of a trustee for realisation and distribution among the creditors and there are many similarities in the way in which it operates in the two jurisdictions. There are also a number of differences, however, some of which are significant in the context of the reform process. Sequestration is initiated by an application to the court, which may be made by the debtor *or* by a creditor or creditors, and is known as voluntary surrender in the former case and compulsory sequestration in the latter. In each case, it is a requirement that the sequestration be to the (pecuniary) advantage of creditors, although that requirement is easier to satisfy in relation to compulsory sequestrations.[31] Following sequestration, the Master may appoint a provisional trustee to administer the sequestration process until a trustee is elected by the creditors or appointed by the Master. Sequestration vests the assets of the debtor, with limited exceptions, initially in the Master, then in the provisional trustee, if any, and ultimately in the trustee; it also vests the assets of the solvent spouse of an individual debtor not married in community of property initially in the Master, then in the provisional trustee, if any, and ultimately in the trustee,[32] although the solvent spouse may ultimately secure the release of defined categories of property. The trustee sells or otherwise realises the assets and distributes them to the creditors in accordance with their rights, but where there are insufficient funds to pay the costs of the sequestration, certain creditors are liable to make a contribution to the costs of the sequestration. There is no provision for the equivalent of the commissioners in a Scottish sequestration, but the creditors as a whole may give the trustee directions in connection with the administration of the estate. The provisions for rehabilitation of the debtor are, on the whole, much stricter than the equivalent Scottish provisions on discharge. Unless the court orders otherwise

[31] In a compulsory sequestration, the court must be satisfied that *there is reason to believe* that sequestration will be to the advantage of creditors whereas in a voluntary surrender, the court must be satisfied that sequestration *will* be to the advantage of creditors. This has led to the problem of so-called friendly sequestrations, discussed further below. In Scotland, there is no such requirement, although there are some additional requirements to be satisfied where an individual debtor petitions for his or her own sequestration.

[32] Where spouses are married in community of property, they have a joint estate and it is that joint estate which is sequestrated. The vesting of the property of the solvent spouse of an insolvent debtor has been declared constitutional: see *Harksen v Lane NO & others*, 1998 (1) SA 300 (CC).

on the application of an interested person, the debtor will be automatically rehabilitated after a period of 10 years, but may apply to be rehabilitated by the court within that period in the circumstances set out in Act 24 of 1936. Rehabilitation has the effect of bringing the sequestration to an end and relieving the debtor of liability for pre-sequestration debts with limited exceptions. There is provision for the debtor to conclude a compromise with creditors after sequestration under section 119 of Act 24 of 1936. Unlike a discharge on composition in Scots law, such a composition does not result in automatic rehabilitation of the debtor, but where the composition is for more than 50 cents in the rand, the debtor is entitled to apply to the court for rehabilitation immediately; otherwise, the debtor may apply for rehabilitation in the normal way where it is wished to be rehabilitated within the 10 year period.

A section 311 compromise is the equivalent of a section 425 arrangement in Scotland and the respective provisions are essentially to the same effect. An application is made to the court for authority to call the relevant meetings. Where that authority is granted, the relevant meetings are called and may accept the proposed compromise or arrangement by the requisite majority.[33] The matter is then referred back to the court, whose sanction is also required; where this is granted, the compromise or arrangement takes effect and binds all members/creditors/classes of members and/or creditors with whom it was concluded and is thereafter implemented accordingly.

Judicial management is a rescue procedure available to companies that are in financial trouble but have the potential to recover. Introduced in 1926, it seems, however, to be regarded as being appropriate only in exceptional circumstances, which limits its utility as a corporate rescue procedure of general application. It is initiated by an application to the court, which may grant an order for judicial management where the company, because of mismanagement or any other cause, is unable to pay its debts or is probably unable to pay its obligations and has not become, or has been prevented from becoming, a successful concern; there is a reasonable probability that, if placed under judicial management, the company will be enabled to pay its debts or meet its obligations or become a successful concern; and it appears just and equitable to grant it. Alternatively, it may be ordered by the court *ex proprio motu* in an application for winding-up of the company. A provisional order providing, *inter alia*, for the company to be managed by a provisional judicial manager, appointed by the Master, and for the existing managerial staff to be divested of the management of the company will be granted initially; it may also stay legal proceedings and process in execution against the company without leave of the court. A report is prepared by the provisional judicial manager and meetings of the company's creditors, members and debenture-holders are convened to consider it and recommend whether the company should be placed finally under judicial management and, if so, to nominate the judicial manager(s). The matter then comes back before the court which will decide whether to make a final order. Where an order is granted, it will vest the management of the company in the judicial manager and set out any directions considered necessary for the management of the company and any incidental matters. The judicial manager will thereafter manage the company in accordance with the court's directions and must apply monies received to the costs of the judicial management and, unless the creditors have resolved otherwise, the claims of pre-judicial management creditors. Judicial management may be brought to an end by the court where it appears either that its purpose has been fulfilled or that it is undesirable for any reason that the order should remain in force.

[33] A majority in number representing three-fourths in value of those present and voting in person or by proxy.

Winding up (or liquidation) is the company equivalent of sequestration and there are many similarities in the way in which it operates in the two jurisdictions. The assets of the company or close corporation are placed in the hands of a liquidator to be realised and distributed to the creditors and, in a solvent winding up, the members and the company or close corporation is then dissolved. There are two types of winding up: voluntary winding up and winding up by the court (also referred to as compulsory winding up). Voluntary winding up is initiated by resolution of the members and there are two types: members' voluntary winding up, where the company or close corporation is able to pay its debts in full, and creditors' voluntary winding up, where the company or close corporation is unable to pay its debts. There is provision for the conversion of a voluntary winding up into a winding up by the court. Winding up by the court is initiated by application to the appropriate court. Following the making of a winding up order or the registration of a resolution for voluntary winding up of a company, the Master may appoint a provisional liquidator and must ultimately appoint a liquidator; following the making of a winding up order or the registration of a resolution for voluntary winding up of a close corporation, the Master must appoint a liquidator. The liquidator gathers in the assets of the company or close corporation, sells or otherwise realises them and applies the proceeds to the costs of the winding up and the claims of the creditors and, in a solvent winding up, distributes any balance to the members. There is no provision for the equivalent of a liquidation committee in a Scottish liquidation, but the liquidator must have regard to any directions given by resolution of the creditors or members and many of the liquidator's powers may only be exercised with the authority of the creditors, members or the Master. On completion of the winding up process, the company or close corporation is dissolved. In the case of a close corporation, there is provision for a composition at any time after the commencement of the winding up of a close corporation unable to pay its debts. Such a composition may provide for the winding up to be set aside by the court.

## REFORMING THE LAW

*Scotland*

Insolvency law in Scotland was subject to major reform in the mid-1980s following the publication in 1982 of both the *Scottish Law Commission Report* and the *Cork Report*. The *Cork Report* was concerned with bankruptcy and company insolvency law in England and Wales, but its recommendations in relation to company insolvency law were of equal relevance to Scotland since, as noted above, the legislation relating to corporate insolvency is mostly common to both jurisdictions and the law is substantially the same in both jurisdictions. The *Scottish Law Commission Report* was concerned mainly with bankruptcy law: the Scottish Law Commission felt that it was inappropriate to embark on a full-scale examination of the law relating to liquidation following the appointment of the Cork Committee, although it did make some recommendations in relation to liquidation in areas where it felt that substantially the same rules should apply in both sequestration and liquidation.[34] These two reports ultimately resulted in the passing of the B(S)A 1985 and the IA 1986 and associated legislation (including the Company Directors Disqualification Act 1986 ("CDDA 1986")) which radically reformed Scottish bankruptcy and company insolvency law

[34] Scottish Law Commission Report, para 1.6.

respectively,[35] but the law has continued to evolve and a variety of further reforms of a more or less radical nature to both bankruptcy and company insolvency law have been implemented or are in contemplation.

Turning first to company insolvency law, the introduction by the IA 1986 of specific company rescue procedures in the form of CVAs and administration was one of the most radical and significant of the changes introduced by that Act, but consultation on further reform to further encourage the development of what is generally referred to as the "rescue culture" began as far back as the early 1990s. A consultative document on CVAs and administration orders identifying barriers to use and seeking views on proposed reforms was issued in 1993.[36] This was followed in 1995 by a further consultative document seeking views on revised proposals for a new CVA procedure, including the introduction of a moratorium,[37] but no legislative action resulted. Then in 1999, having announced its intention to legislate for the introduction of a moratorium procedure for CVAs, the government instituted a general review of company rescue and business reconstruction mechanisms with a remit to review

> aspects of company and insolvency law and practice in the United Kingdom and elsewhere relating to the opportunities for, and means by which, businesses can resolve short to medium term financial difficulties, so as to preserve maximum economic value; and to make recommendations.[38]

The review was specifically tasked to focus on, *inter alia*, the further development of the rescue culture; a re-assessment of the relative rights and remedies of secured and unsecured creditors, including the Crown as a preferential creditor; and the duties of directors of companies experiencing financial difficulties.[39] At the same time, it was announced that a further consultation paper on the effect that insolvency, particularly bankruptcy, law and practice has on enterprise by contributing to the stigma of failure and inhibiting those who would wish to try again in business was planned for the near future[40] and a consultation paper on possible reform of bankruptcy law in England and Wales was duly published in early 2000.[41] The review of company rescue and business reconstruction mechanisms reported in May 2000,[42] by which time the IA 2000 containing, *inter alia*, important provisions for reform of CVAs including the introduction of a moratorium for that procedure,[43] was already on its way through

[35] The IA 1986 also radically reformed bankruptcy law in England and Wales, which, as noted above, had also been within the remit of the Cork Report.

[36] The Insolvency Service, *Company Voluntary Arrangements and Administration Orders: A Consultative Document*, October 1993.

[37] The Insolvency Service, *Revised Proposals for a New Company Voluntary Arrangement Procedure: A Consultative Document*, April 1995.

[38] The Insolvency Service, *A Review of Company Rescue and Business Reconstruction Mechanisms*, 1999, para 1(a).

[39] *Ibid*, para 1(b).

[40] *Ibid*, para 1(e).

[41] The Insolvency Service, *Bankruptcy: A Fresh Start*, 2000. The consultation paper was confined to possible reform of the law in England and Wales because by then (the majority of) bankruptcy law in Scotland was devolved under the Scotland Act 1998, but the progress of bankruptcy reform in England and Wales is discussed here because, as discussed below, it has had an influence on prospective reform to bankruptcy law in Scotland.

[42] The Insolvency Service, *A Review of Company Rescue and Business Reconstruction Mechanisms: Report by the Review Group*, May 2000.

[43] In the event, as noted above, these provisions did not come into effect until January 2003: see the Insolvency Act 2000 (Commencement No 3 and Transitional Provisions) Order 2002, SI 2002/2711. Other important provisions in the IA 2000 included reforms to the provisions on disqualification of directors and extension of the effect of the moratorium and interim moratorium in administration to a landlord's exercise of a right of forfeiture by peaceable re-entry in England and Wales. Equivalent provision for the extension of the effect of the moratorium and interim moratorium in administration to a landlord's exercise of a right of irritancy in Scotland was subsequently made by the EA 2002.

parliament.[44] A White Paper, *Productivity and Enterprise: Insolvency – A Second Chance*[45] ("the White Paper") was then published in July 2001 setting out proposals for further reform of company insolvency law in the UK and reform of bankruptcy law in England and Wales. The aim of the proposed reforms was stated as being:

> [to] modernise the framework of personal and corporate insolvency. . .encourage responsible risk-taking, facilitate the rescue of viable businesses and provide certainty and fairness to creditors and other stakeholders. . . address the fear of failure and reduce the stigma of bankruptcy . . .[46]

For company insolvency law, this was to be achieved by placing restrictions on the use of administrative receivership resulting in increased use of administration, which was to be streamlined to make it more effective and accessible, and by abolishing Crown preference and taking steps to ensure that the benefit of that abolition would inure to unsecured creditors; for personal insolvency law, it was to be achieved by reducing the period within which a bankrupt receives an automatic discharge and removing many of the restrictions applying automatically on bankruptcy but at the same time, in order to deter abuse of the liberalised regime, introducing bankruptcy restriction orders ("BROs") similar to disqualification orders under the CDDA 1986 and designed to provide for the continuance of bankruptcy restrictions against what was described as a "small minority of culpable bankrupts". The abolition of Crown preference would also apply in bankruptcy. The response to the proposals was generally positive, but concerns were raised about a number of aspects of both the company insolvency and bankruptcy proposals[47] and the government published two further papers setting out its response to the concerns raised by the consultation in January and March 2002.[48] The Enterprise Bill was published on 26 March 2002, the same day as the second paper, and after a somewhat stormy passage through parliament, became the EA 2002 on 7 November 2002. As noted above, the company insolvency law provisions of the EA 2002 came into force on 15 September 2003 while the bankruptcy law reforms in England and Wales came into force on 1 April 2004.

Turning to bankruptcy law, important changes to bankruptcy law in Scotland were made by the Bankruptcy (Scotland) Act 1993 ("B(S)A 1993"), including the extension of the role of the AIB to encompass for the first time the actual administration of sequestrated estates.[49] A consultation paper seeking views on "how to resolve problems in the [B(S)A 1985], which have led to some debtors with intractable debt problems being unable to obtain the relief of sequestration" was issued in July 1997[50] and followed by a further consultation paper seeking views on a number of proposed amendments to the legislation in this respect and on aspects of protected trust deeds for creditors, the payment of dividends in sequestration and the qualifying debt level

[44] This attracted a good deal of criticism at the time, on the basis that it would have been more sensible to await the outcome of the review and then enact all necessary changes in one piece of legislation. In the event, the additional reforms were contained in the EA 2002.

[45] Cm 5234.

[46] White Paper, Executive Summary.

[47] See Summary of Responses to the White Paper "*Productivity and Enterprise – Insolvency: A Second Chance*" published at http://www.insolvency.gov.uk/wpsum.htm

[48] *An update on the Corporate Insolvency Proposals* was published on 14 January 2002 and *An update on the Bankruptcy Proposals* was published on 26 March 2002; both of these can found at http://www.insolvency.gov.uk/

[49] See above.

[50] The Scottish Office, *Apparent Insolvency: A Consultation Paper on Amending The Bankruptcy (Scotland) Act 1985*, July 1997. The consultation concerned the definition of apparent insolvency, which is a pre-requisite for sequestration in most cases.

for sequestration in July 1998,[51] but no legislation resulted. Other important changes subsequently took place – provisions exempting most pensions from vesting in the debtor's trustee in sequestration came into force in May 2000,[52] complementary provisions allowing the trustee to recover excessive and unfair pension contributions made by the debtor prior to sequestration came into force in April 2002[53] and the list of property exempt from diligence, and hence sequestration, was expanded in 2000[54] and again in 2002[55] – but no further structural reforms to sequestration itself.

Other related reform has taken place in the form of the introduction of the DAS. A debt arrangement scheme had been proposed by the SLC as far back as 1980[56] and recommended in a report in 1985[57] but never implemented. Calls for the introduction of some form of debt arrangement scheme were renewed, however, during the debates surrounding the Abolition of Poindings and Warrant Sales Act 2000 ("APWSA 2000")[58] and in the reviews of both poinding and warrant sale in particular and diligence generally to which it gave rise,[59] and ultimately resulted in the introduction of the DAS by the DAA(S)A 2002.[60]

[51] The Scottish Office, *The Bankruptcy (Scotland) Act 1985, A Consultation Follow-Up: Protected Trust Deeds and Other Issues*, July 1998.

[52] Welfare Reform and Pensions Act 1999, ss 11 and 12, as applied to Scotland by s 13, brought into force by the Welfare Reform and Pensions Act 1999 (Commencement No 7) Order 2000, SI 2000/1382. Prior to the coming into force of these provisions, a debtor's rights under an occupational pension scheme would not generally have formed part of his or her estate in any event because the pension fund in such a case is not normally regarded as the property of the debtor (although he or she would have been potentially liable to make a contribution from any pension income received during the sequestration under the provisions relating to contributions from income), but a retirement annuity contract or personal pension policy would have formed part of the debtor's estate with the result that all monies payable in respect of the pension became payable to the trustee *whenever paid*: for further detail, see McKenzie Skene, *Insolvency Law in Scotland*, (Edinburgh: T&T Clark, 1999) 166–7.

[53] Welfare Reform and Pensions Act 1999, s 16, brought into force by the Welfare Reform and Pensions Act 1999 (Commencement No 13) Order 2002, SI 2000/153. Similar provisions had previously been enacted by the Pensions Act 1995, but had never come in force.

[54] See the Debtors (Scotland) Act 1987 (Amendment) Regulations 2000, SI 2000/189.

[55] See the DAA(S)A 2002, Sch 3, para 15(3).

[56] Scottish Law Commission *Fourth Memorandum on Diligence: Debt Arrangement Schemes*, Scot. Law Com. Consultative Memorandum No 50 (1980).

[57] Scottish Law Commission *Report on Diligence and Debtor Protection*, Scot. Law Com. No 95 (1985). The report contained detailed provisions for the scheme, some, but not all, of which have been adopted in the new DAS.

[58] That Act began life as a private member's bill providing for the abolition of the diligence of poinding and warrant sale, being the diligence used to attach moveable property of a debtor in that debtor's own possession in satisfaction of the attaching creditor's debt. The diligence had become unpopular due to its extensive use by local authorities in enforcing payment of certain taxes due to these local authorities by individual debtors.

[59] The specific issue of the abolition of poinding and warrant sale was referred to the SLC by the Scottish Executive in 1999 in response to the lodging of the proposal for the private member's bill which ultimately became the APWSA 2000. The SLC duly published a discussion paper *Poinding and Sale: Effective Enforcement and Debtor Protection*, Scot Law Com Discussion Paper No 110, in November 1999 in which, *inter alia*, it re-opened the question of the introduction of a debt arrangement scheme. Its subsequent *Report on Poinding and Warrant Sale*, Scot Law Com No 177, published in April 2000, once again recommended the introduction of such a scheme, although this time it did not make specific recommendations as to how such a scheme should be constructed, a matter which it recommended should be the product of consultation with both debtor and creditor interests. In addition, the Scottish Minister for Justice had also established a working group charged with bringing forward proposals for an acceptable alternative form of diligence for attaching moveable property in the hands of a debtor and announced a review of the law of diligence generally in June 2000. The report of the working group, *Striking the balance – a new approach to debt management*, published in June 2001, also recommended the introduction of a statutory debt arrangement scheme. In April 2002, following completion of its general review of the law of diligence, the Scottish Executive published a consultation paper entitled *Enforcement of Civil Obligations in Scotland* which contained detailed proposals for a statutory debt arrangement scheme.

[60] As noted above, the DAS came into force on 30 November 2004. The DAA(S)A 2002 was enacted before the outcome of the above-mentioned consultation on enforcement of civil obligations was available, with the result that the legislation itself set out only the basic framework of the scheme, leaving the details to be provided for in regulations which were ultimately enacted in the form of the Debt Arrangement Scheme (Scotland) Regulations 2004, SS/468 as amended. For a discussion of the history of the DAS and the provisions contained in the DAA(S)A 2002 itself, see McKenzie Skene, "Dealing with multiple debt – an examination of the proposals for a debt arrangement scheme in Scotland" [2002] *Insol L* 212; for a discussion of the details of the scheme based on a draft of the regulations see McKenzie Skene, "The Debt Arrangement Scheme", 2003 SLT (News) 289; and for a discussion of the final version of the scheme as ultimately brought into force, see McKenzie Skene "The Debt Arrangement Scheme Goes Live" 2004 SLT (News) 237.

Further bankruptcy reform is now in prospect. In November 2003, the Scottish Executive published a consultation paper entitled *Personal Bankruptcy Reform in Scotland: A Modern Approach* ("the First Consultation Paper") seeking views on a number of proposed reforms to bankruptcy law in Scotland. In July 2004, this was followed by a further consultation and draft bill, *Modernising bankruptcy and diligence in Scotland: Draft Bill and Consultation* ("the Second Consultation Paper") setting out firm proposals on which it is intended to legislate and proposals for further consultation and consideration.[61] As in England and Wales, reform of bankruptcy law in Scotland is seen as an important step in building a modern and prosperous Scotland.[62] The main aims of the bill as set out in the Second Consultation Paper include modernising the law of personal bankruptcy and diligence in Scotland *and* striking a balance between the differing needs of creditors and debtors and encouraging responsible risk taking[63] and the Second Consultation Paper went on to state in relation to the proposed reforms to bankruptcy law that:

> [t]he reform proposals seek to strike a balance between encouraging people to get on with their lives and start again after bankruptcy with the need to protect public and business communities from reckless spending behaviour. Their overall thrust is to encourage personal and business restart whilst upholding a "can pay, should pay" principle.[64]

The First Consultation Paper identified two drivers for change behind the proposals for reform: first, the importance of having an integrated debt management framework in Scotland within which the debt management tools available work together to form a comprehensive package of solutions[65] and second, what it described as "developments", being the introduction of the DAS and the consequent need to consider its fit with sequestration,[66] the need to consider whether action is still required in relation to the issues raised in the earlier consultations on apparent insolvency and protected trust deeds (referred to above),[67] and the significant changes to the law of personal bankruptcy in England and Wales, in particular the reduction of the period of bankruptcy to one year, brought about by the EA 2002.[68] It sought views on a wide range of possible reforms, including:

(i) The introduction of a one-year bankruptcy period for all debtors similar to that introduced in England and Wales, with certain limited exceptions.[69] Two reasons were

[61] Second Consultation Paper, para 3.3.

[62] First Consultation Paper, para 1.6. Although there is no direct reference to it, this approach is consistent with that being developed at a European level: see the study commissioned by the European Commission *Bankruptcy and a fresh start: Stigma on failure and legal consequences of bankruptcy* (Brussels, 2002): http://europa.eu.int/comm./enterprise/entrepreneurship/support_measures/failure_bankruptcy/bankruptcy.htm and the European Commission's *Best Project on Restructuring, Bankruptcy and a Fresh Start: Final Report of the Expert Group* (September 2003): http://europa.eu.int/comm./enterprise/entrepreneurship/support_measures/failure_bankruptcy/index.htm

[63] Second Consultation Paper, para 1.1.

[64] *Ibid*, para 3.1. The First Consultation Paper described the proposals as intended to "reduce the stigma of bankruptcy and encourage responsible risk taking while providing a robust and effective regime to protect the public and business community from the small minority of bankrupts who have acted in a fraudulent or culpable manner": First Consultation Paper, para 1.7.

[65] First Consultation Paper, para 3.1.

[66] *Ibid*, para 3.3. This could equally well be seen as an aspect of the first driver for change.

[67] *Ibid*, para 3.4.

[68] *Ibid*, para 3.5.

[69] The possible exceptions discussed in the First Consultation Paper were bankrupts who failed to co-operate with their trustee, in respect of whom it was proposed that the existing option to defer discharge should be retained, and repeat bankrupts, in respect of whom a five year period of bankruptcy would be imposed. Repeat bankrupts were defined as those bankrupt for a third or subsequent time *or* for a second time where a bankruptcy restriction order had been made in relation to the first bankruptcy (bankruptcy restriction orders are discussed further below). Views were also sought on any other categories of debtor for whom the proposed new reduced period of bankruptcy would be inappropriate. On consultation, there was widespread support for a longer period of bankruptcy for repeat bankrupts and for those who had failed to co-operate with their trustee or were otherwise culpable (Second Consultation Paper, para 5.16) and it is

given for this proposal. The first was a perception that the current three-year bankruptcy period could be harsh for small business bankrupts and the Scottish Executive wished to encourage those failing honestly to try again.[70] This is the same reasoning as that behind the reforms in England and Wales. The second was the necessity of ensuring a level playing field for business throughout the UK, consistency in bankruptcy periods being part of this.[71] The reasons for the proposed change were thus essentially enterprise-oriented but, as noted, it was intended to apply to all bankrupts whether in business or not.[72] The introduction of a one-year bankruptcy period was the proposal on which the views of consultees were most evenly divided,[73] but the Scottish Executive has nonetheless decided to press ahead with this reform, believing that safeguards to be put in place (discussed further below) will meet any concerns.[74]

(ii) Changes to the qualifying level of debt required for sequestration. At present, the qualifying level of debt for both debtor and creditor petitions is £1,500 and the First Consultation Paper sought views on whether this level should be altered for debtors or creditors or both. The qualifying level of debt is important because it affects access to the sequestration process and the First Consultation Paper acknowledged the need to strike an appropriate balance between protecting debtors from being subjected to sequestration for small sums and the needs of creditors with respect to enforcement.[75] Most consultees favoured the limit remaining unchanged and the Scottish Executive agreed; the limit will therefore remain unchanged for the present.[76]

(iii) If a one-year bankruptcy period were introduced, changes to the current provisions on contributions from income by the debtor to allow these to continue beyond discharge in a way similar to that provided for in England and Wales. At present, any income contributions by the debtor continue only until discharge, thus any reduction in the bankruptcy period would lead to a reduction in income contributions absent an extension of the period for contributions beyond discharge. The issue is again one of balancing the interests of the debtor in obtaining a fresh start as quickly as possible and the creditors in receiving a reasonable contribution from the debtor.[77] Approximately two-thirds of consultees were in favour of extending the period within which income contributions should be exigible beyond discharge,[78] although there was no discernible consensus on whether that period should be fixed or variable.[79] It is therefore now intended that the period during which contributions will be exigible will extend beyond the debtor's discharge, but that payments should be made for a maximum period of three years.[80] It is notable, however, that notwithstanding that the matter was raised by a number of consultees responding to the First Consultation Paper, there is no corresponding proposal to allow after-acquired property to continue

intended that a five year bankruptcy period for repeat bankrupts will be provided for in the Bill at introduction (Second Consultation Paper, para 5.17). With respect to the other categories, it seems to be intended that these will be dealt with by way of deferral or a bankruptcy restriction order (Second Consultation Paper, para 5.22).

[70] *Ibid*, para 4.3.

[71] *Ibid*, para 4.4.

[72] Two reasons were given for this approach: first, the difficulty of distinguishing business bankruptcies from non-business bankruptcies and second, the creation of a two-tier system which would provide minimal benefits, complicate aspects of the sequestration process and would not be consistent with the idea of reducing stigma: see *ibid*, para 4.5.

[73] Second Consultation Paper, para 5.4.

[74] *Ibid*, para 5.8, 5.10.

[75] First Consultation Paper, para 5.9.

[76] Second Consultation Paper, para 5.67.

[77] See *ibid*, para 5.24.

[78] *Ibid*, prior to para 5.23.

[79] *Ibid*, para 5.27.

[80] *Ibid*, paras 5.24, 5.28.

to be caught beyond discharge,[81] with the result that a reduced bankruptcy period will result in a potential loss to creditors from this source. These proposed changes will have an impact on the relative attractiveness of sequestration and other alternatives open to the debtor, for example the DAS, where there is no maximum period for payments.[82]

(iv) Changes to the provisions on apparent insolvency designed to make access to sequestration by the debtor easier. Apparent insolvency, defined by the B(S)A 1985, section 7, is a pre-requisite for a creditor petition for sequestration and one of two alternative means of satisfying one of the pre-requisites for a debtor petition for sequestration where there is no concurrence by a qualified creditor.[83] Establishing apparent insolvency for this purpose can prove problematic for some debtors, however, and although some changes to the concept had previously been made, the First Consultation Paper sought views on whether further changes were required. Again, this affects the question of access to the sequestration process. A large majority of consultees were in favour of further changes, and it is therefore intended to introduce appropriate provisions although further views are being sought on the precise nature of these.[84] A Working Group on Debt Relief has also been established which is considering this issue, *inter alia*.[85]

(v) Integration of bankruptcy and the DAS. The inter-relationship of the DAS and sequestration, and indeed the other options open to a debtor, is critical in the context of producing a coherent system for dealing with debtors. The First Consultation Paper suggested that it would be better for a debtor who would be suitable for the DAS to be diverted into that scheme and views were sought on whether, and if so how, this should be done.[86] Consultees generally supported the idea of encouraging suitable debtors to enter a DAS, but there was less consensus on an appropriate mechanism.[87] It is not therefore proposed to impose any conditions which would require a debtor to divert to the DAS and the matter will therefore remain one of debtor choice for the moment.[88]

(vi) Changes to the range of debt management tools available. The First Consultation Paper emphasised the need to provide a comprehensive debt management framework with appropriate solutions for all debtors[89] and views were sought on whether the existing and proposed solutions would achieve this. The primary concern here is those debtors with little or no assets or income, for whom none of the existing solutions may be accessible or appropriate,[90] and while not attracted to creating a separate scheme for such debtors, the Scottish Executive acknowledged that the matter required further consideration and established the Working Group on Debt Relief

[81] At present, subject to certain provisos, property acquired between the date of sequestration and the date of the debtor's discharge currently falls into the sequestration.

[82] This is discussed further below.

[83] In practice, most debtors seeking to petition for their own sequestration without the concurrence of a qualified creditor would be seeking to rely on their apparent insolvency to satisfy this pre-requisite.

[84] Second Consultation Paper, paras 7.1-7.7.

[85] See *ibid*, paras 7.3, 7.10. The remit of the Working Group on Debt Relief can be found at http://www.scotland.gov.uk/Topics/Justice/Civil/17868/BackgroundRemit.

[86] First Consultation Paper, para 6.10.

[87] Second Consultation Paper, para 5.69.

[88] *Ibid.*

[89] First Consultation Paper, para 6.24.

[90] Second Consultation Paper, para 7.8.

referred to above to give further consideration to this issue *inter alia*.[91] Although not mentioned in the Second Consultation Paper, another area of concern here might be the lack of a rescue-oriented procedure for non-company trading debtors.[92]

(vii) The introduction of time limits for dealing with the bankrupt's home and certain other assets. The First Consultation Paper sought views on the introduction of a three-year time limit for dealing with the bankrupt's home, subject to extension in appropriate circumstances.[93] The majority of consultees was in favour of this proposal and it is therefore intended that the debtor's home should revert to the debtor if the trustee has not commenced dealing with it within three years of the date of the bankruptcy or the date of discovery of its existence.[94] The First Consultation Paper also sought views on whether the trustee should only be able to "claim" inheritances and other legal rights where these materialised within a particular period.[95] Again, the majority of consultees was in favour of such a provision and it is therefore intended that rights to inheritances and legal rights that have not materialised prior to the debtor's discharge will revert to the debtor on discharge.[96] Rights arising under an irrevocable deed such as an established trust will not, however, revert to the debtor under this provision.[97]

(viii) Improved monitoring and transparency of protected trust deeds. The First Consultation Paper identified a number of concerns previously expressed about protected trust deeds, including a concern that not all protected trust deeds are for the benefit of creditors, a number of them seemingly being designed solely to save the debtor from the more serious consequences of sequestration.[98] It therefore sought views on a number of changes designed to improve the monitoring and transparency of protected trust deeds.[99] The main proposals were widely supported by consultees and it is therefore intended to implement them accordingly.[100]

(ix) If a one-year bankruptcy period were introduced, the introduction of a bankruptcy restriction order regime similar to that introduced in England and Wales. Both the First and Second Consultation Paper acknowledge the importance of striking a balance between encouraging people to get on with their lives and start again after bankruptcy on the one hand and the need to protect the public and businesses from reckless spending behaviour on the other[101] and recognised the need for appropriate safeguards to accompany a reduced bankruptcy period.[102] The introduction of a bankruptcy restriction order regime is seen as a means of continuing controls on, and

[91] See *ibid*, para 7.10.

[92] This is discussed further below.

[93] First Consultation Paper, paras 7.3, 7.4. Although not mentioned in the First Consultation Paper, a similar provision was introduced in England and Wales by the EA 2002.

[94] Second Consultation Paper, paras 5.30, 5.31.

[95] See *ibid*, paras 7.7 to 7.10. The term "claim" was somewhat misleading, since it was proposed that the rights in question would continue to vest in the trustee as at present but the trustee would lose the right to them after a given period.

[96] *Ibid*, paras 5.33, 5.35.

[97] *Ibid*.

[98] First Consultation Paper, para 8.5.

[99] *Ibid*, para 8.7.

[100] Second Consultation Paper, para 6.6. It is intended to do this by regulations to be introduced at the same time as the bill, rather than primary legislation.

[101] See, *eg*, First Consultation Paper, para 4.4; Second Consultation Paper, para 3.1.

[102] First Consultation Paper, paras 4.4, 4.6; Second Consultation Paper, para 5.8.

thus protecting the public and businesses from, bankrupts who have acted in a potentially fraudulent or culpable manner.[103] The regime is intended to cover a wide range of "culpable" conduct by both trading and consumer debtors.[104]

(x) A review of current bankruptcy restrictions. The First Consultation Paper proposed the development of a more focused approach to the issue of bankruptcy restrictions with a view to removing outdated restrictions conferring stigma without offering protection to the public while retaining appropriate restrictions and considering what restrictions should flow from a bankruptcy restriction order regime if introduced.[105] The Second Consultation Paper adopts a radical approach to this issue. It intends that the only disqualifications flowing automatically from bankruptcy itself will be disqualification from acting as a company director and disqualification from acting as an insolvency practitioner,[106] while any other disqualifications will be linked to the bankruptcy restriction order regime if appropriate.[107] One particular aspect of bankruptcy restrictions is credit restrictions on bankrupts. At present, it is an offence for an undischarged bankrupt to incur credit in excess of £250 without disclosing his or her status to the lender.[108] The First Consultation Paper sought views on whether this limit, which has remained unchanged for some time, should be increased. Most consultees were in favour of this and it is therefore intended to increase this limit to £500.[109] The First Consultation Paper also sought views on whether there should be a total credit limit applied to a bankrupt and, if so, the appropriate level.[110] Again, the majority of consultees was in favour of this and it is therefore intended to introduce a total credit limit which would require an undischarged bankrupt to inform a lender of his or her status irrespective of the amount of credit sought in the particular case if that limit had been reached or would be exceeded by the borrowing.[111] It is provisionally intended that this total credit limit should be set at £1,000, but further views on the appropriate level are sought.[112]

(xi) Changes to streamline procedures. The First Consultation Paper set out a number of proposals for streamlining procedures. These included:

> (a) Consolidating all bankruptcy proceedings in the sheriff court. This was supported by the majority of consultees and it is intended to implement this proposal.[113] A proposal that debtor petitions should be processed by the AIB rather than the courts was also generally supported by consultees and it is also intended to implement this proposal.[114]

[103] First Consultation Paper, paras 9.5, 9.6; Second Consultation Paper, paras 3.2, 5.8, 5.39, 5.40.

[104] Second Consultation Paper, paras 5.36, 5.39. The draft bill contains a non-exhaustive list of conduct to be taken into account in this context.

[105] First Consultation Paper, para 9.11. It also sought views on some of the existing restrictions.

[106] Second Consultation Paper, para 5.46.

[107] *Ibid*, paras 5.45–5.52. Some specific disqualifications which it is intended should be linked to the bankruptcy restriction order regime are identified in the Second Consultation Paper itself and provision is made for further changes in this area to be made by secondary legislation to allow review of the whole area as appropriate: Second Consultation Paper, para 5.50.

[108] B(S)A 1985, s 67(9).

[109] Second Consultation Paper, para 5.53. The corresponding limit in England and Wales has also been increased to the same figure.

[110] First Consultation Paper, para 9.15.

[111] Second Consultation Paper, paras 7.13, 7.14.

[112] *Ibid*, para 7.15. It should be noted that it is intended that credit incurred in respect of what is described as essential utilities such as gas, electricity, water and council tax should be excluded from the calculation of both the single-lender credit limit and the total credit limit: see para 5.54.

[113] First Consultation Paper, paras 5.62–5.64. Reduction and suspension of sequestration and certain matters relating to incoming foreign judgements will, however, continue to be dealt with by the Court of Session: para 5.63.

[114] *Ibid*, para 5.56.

(b) Combining the roles of interim and permanent trustee in sequestration. Again, this was supported by the majority of consultees and it is intended to proceed with this proposal.[115]

(c) Streamlining the procedure for judicial composition to minimise court time and expense. Again, this was supported by the majority of consultees and it is intended to proceed with this proposal.[116]

It is anticipated that a bill will be introduced in the next parliamentary session.

*South Africa*

In South Africa, non-company insolvency law had been under review by the SALC for some time prior to the submission of the SALC report in 2000. The report explains that Act 24 of 1936, which replaced the Insolvency Act 32 of 1916 but did not amend it drastically, has been amended more than 20 times but has never been reviewed as a whole. The SALC appointed a project committee, which met 28 times, to assist it with the project; six interim reports were submitted and seven working papers published for comment; and discussion papers with draft bills and explanatory memoranda were published for comment in 1996 and 1999 respectively.[117] The aim of the investigation as stated in the SALC report itself was "to balance and satisfy the needs of the different stakeholders", the major stakeholders being "the commercial community in general and creditors in particular; insolvent debtors; insolvency practitioners and the government", although it was recognised that because of conflicting interests, it is often difficult to strike a fair balance between the different interests.[118] The report discloses that a number of broad principles or guidelines for reform underlie its recommendations: it concludes that, having considered the general principles of a number of legal systems and proposals for reform:

> ... it is clear that similar solutions are used [in different systems of insolvency] because of similar problems experienced worldwide. There is a general need for efficient procedures to deal with assets of the insolvent estate. If liquidation ensues, a fair distribution of the proceeds of the assets must be effected as soon as possible. All systems contain measures to investigate the affairs of the insolvent and to set aside some transactions entered into to the detriment of the general body of creditors. Systems require from debtors to co-operate but afford them an opportunity for a fresh start.[119]

The report goes on to emphasis that effective, speedy and fair procedures are important needs of stakeholders and form the basis of the review[120] and that particularly in difficult economic times it is important that money should be available to generate growth and should not be entangled in tiresome and time consuming processes.[121] This echoes the sentiments expressed in the White Paper.

It is noted in the report that many of the proposed changes contained therein are technical,[122] but that the value of practical or technical reforms should not be underestimated.[123] The main substantive changes proposed in the report include:

[115] *Ibid*, paras 5.72, 5.73. There will still be provision for the appointment of an interim trustee before the award of sequestration is made in appropriate cases.

[116] *Ibid*, para 5.78.

[117] South African Law Commission, Project 63, *Review of the Law of Insolvency*, Discussion Paper 66 (1996) and Discussion Paper 86 (1999) respectively.

[118] SALC Report, para 2.3.

[119] *Ibid*, para 2.1.

[120] *Ibid*, para 2.3.

[121] *Ibid*, para 2.4.

[122] *Ibid*, Summary, p 6.

[123] *Ibid*, para 2.5.

(i) A new requirement for the liquidator[124] to be a member of a recognised professional body. This was included "[i]n the light of consistent complaints that some liquidators act dishonestly or that they are not competent".[125] This proposal would bring the position closer to that in the UK, where the current provisions on the regulation of IPs were introduced for similar reasons,[126] although the regulatory requirements in the UK are much stricter.[127] A Commission of Enquiry on the Liquidations Industry established in 2004 is, *inter alia*, considering this issue and is due to report at the end of January 2005.

(ii) Changes in the role of the Master. The proposals place more responsibilities on the creditors and reduce the role of the Master,[128] an approach which is consistent with the principle stated later in the report that creditors should accept responsibility for the protection of their own interests.[129] These proposals would make the role of the Master closer to that of the AIB in Scotland in cases where the AIB is not actually administering the sequestration process.

(iii) Changes in the treatment of some creditors, including the abolition of many of the preferent claims currently provided for in Act 24 of 1936.[130] In particular, it is proposed to abolish most of the preferences in favour of State institutions and to retain out of those preferences payable out of the free revenue of the estate only those in favour of employees, contributions to employee funds and claims for arrear maintenance payable in terms of a court order. This proposal would result in a position similar to that in Scotland following the abolition of Crown preference by the EA 2002, except in relation to claims for arrears of maintenance, which are not preferred claims in Scotland.

(iv) Changes in the provisions for setting aside prior transactions. A number of problems had been identified with the existing provisions for setting aside prior transactions and the proposals are designed to make the provisions easier to enforce.[131] It is notable that although every insolvency system has such provisions, they can vary considerably. The main provisions relating to the setting aside of prior transactions in Scotland are essentially the same in both bankruptcy and company insolvency law[132] but differ from those in England and Wales[133] and the South African provisions have both similarities to and differences from the provisions in both of these jurisdictions. It is interesting, then, that one of the proposed changes would make the relevant South African provision more like the equivalent provisions in both Scotland and England and Wales by widening the application of the relevant provision in cases involving

[124] New terminology, including the substitution of the term "liquidator" for "trustee", was adopted in the draft bill to facilitate the incorporation of its provisions into unified legislation dealing with both company and non-company insolvency also under consideration: SALC Report, paras 5.1-5.4. This is discussed further below.

[125] SALC Report, para 4.7.

[126] See the *Cork Report*, Ch 15.

[127] These are outlined above.

[128] SALC Report, para 3.5.

[129] *Ibid*, para 4.3

[130] The equivalent of preferred/preferential claims in Scots law. This is discussed further below.

[131] See SALC Report, para 20.1 *et seq*.

[132] Although there are some provisions which apply only in one or the other due to the differing nature of the debtors involved: bankruptcy law contains provisions for the setting aside of certain orders on divorce and the recovery of excessive or unfair pension contributions which are clearly appropriate only for individual debtors while company insolvency law contains a provision for the avoidance of certain floating charges which is appropriate only for company or LLP debtors since only such debtors may grant floating charges.

[133] This is one of the relatively few areas where there is a substantive difference in company insolvency law in the two jurisdictions.

associates of the debtor,[134] while another of the proposed changes addresses an issue which has also been problematic in the equivalent provision in England and Wales but which does not arise in Scotland.[135]

(v) The introduction of a cap on the exclusion of pension benefits from the insolvent estate and the introduction of provisions for the recovery of certain extraordinary contributions to pension funds for the benefit of creditors. As noted above, it is only recently that legislation excluding most types pension from sequestration was introduced in Scotland and the new Scottish provisions for recovery of excessive and unfair contributions to pensions were seen as a counterbalance to that provision to prevent its possible abuse by the debtor, who would otherwise effectively have been able to put assets beyond the reach of creditors by placing them in a pension fund. It is interesting, therefore, that it has been considered appropriate to introduce a cap on the excluded benefits in South African law which will apply even if the contributions to the pension are not challengeable in this way.

(vi) Relaxation of the existing requirements for a composition following on liquidation and the introduction of new provisions for a binding composition between a debtor and a majority of creditors outwith formal insolvency proceedings. The provisions for the proposed new composition procedure appear in the form of a new section 74X of Act 32 of 1944. Designed to provide a solution for debtors with little or no assets who through no fault of their own are unable to pay their debts and for whom liquidation[136] would not be appropriate,[137] by providing for the composition to be binding on all creditors if accepted by a specified majority of them, the new procedure avoids the disadvantage of a common law compromise where the consent of all creditors is required. The proposals provide for the composition to be supervised by a magistrate and to take place only after investigation into the debtor's affairs. Where an offer of composition is made by a debtor and is not accepted by the relevant majority, and the court is of the view that the debtor is unable to make available substantially more than that offered, provision is made for the debtor to consent to liquidation of his or her assets in terms of Act 24 of 1936 but without being regarded as insolvent or his or her estate as being liquidated when any other Act is being applied. These proposals, however, attracted some criticism,[138] in particular the option for the debtor to consent to liquidation where an offer of compromise is not accepted, which has been described as potentially open to abuse in practice.[139] An interim research report by CACIL reviewing the existing provisions on administration orders under Act 32 of 1944[140] ("the second CACIL Report") published in May 2002 recommends, *inter alia*, that the proposed new composition procedure should be adopted, but subject to clarification of its relationship with the administration order procedure.[141] The recommendations of the second CACIL Report are explored further below.

[134] This is the term used in both Scotland and South Africa – the equivalent term in England and Wales is "connected person".

[135] This is the requirement to establish an intention to prefer in certain cases; this is not a requirement in Scotland.

[136] See note 124 above in relation to terminology.

[137] SALC Report, para 124.1.

[138] See *ibid*, para 124.1 *et seq.*

[139] See Boraine, A and Roestoff, M "Fresh Start Procedures for Consumer Debtors in South African Bankruptcy Law", 2002 IIR 1.

[140] CACIL, *Interim Research Report on the Review of Administration Orders in Terms of The Magistrates' Courts Act 32 of 1944*, May 2002.

[141] *Ibid.*

One important, and controversial, area where the SALC Report recommends no change is the requirement for liquidation to be to the advantage of creditors. The report notes that:

> The requirement of "advantage to creditors" does not apply in the case of companies and is not common in other legal systems. The requirement has been criticised because of the difficulties to prove an advantage of creditors and the result that a debtor cannot escape from the quagmire of debts if he or she does not own unencumbered assets of sufficient value.[142]

The requirement of advantage to creditors has given rise to the particular problem of "friendly sequestrations". As noted above, it applies in both voluntary surrender and compulsory sequestration, but is easier to establish in relation to the latter. Thus, a debtor who wishes to access the sequestration procedure but who would not be able to meet the more stringent requirements for a voluntary surrender may seek to arrange for a "friendly" creditor to institute a compulsory sequestration which is more likely to be granted due to its less stringent requirements. This practice has been seen as an abuse of the process of the court and has given rise to complex case law as the courts have sought to impose special requirements in friendly sequestrations to prevent this,[143] which are acknowledged as problematic in the SALC Report.[144] Nevertheless, the report goes on to say:

> It is submitted that it is unacceptable to use the expensive procedure of liquidation by the court in cases where the value of the assets is insufficient to ensure a benefit to creditors. Schedule 4 of the Bill proposes a procedure which will assist debtors to arrive at a composition with creditors before liquidation. Provision is also made for a less expensive liquidation procedure if a composition cannot be reached with the requisite majority of creditors notwithstanding a reasonable offer by the debtor.[145]

Thus, the solution is seen as providing effective alternatives to formal insolvency in such cases rather than changing the requirements of the formal insolvency procedure itself. This is discussed further below. It is important to note, however, that the SALC Report explicitly accepts that a debtor may become insolvent through no fault of his or her own and that such a debtor should be given the opportunity to make a fresh start. It is acknowledged that creditors sometimes contribute towards insolvencies by giving credit to those who cannot repay it and that a balance must be struck between the rights of creditors and giving a debtor an opportunity to make a fresh start.[146]

The report states that it is "an important first step to modernise the law of insolvency but is clearly not a final assessment of the subject"[147] and indeed further developments in the reform process were already underway when the report was submitted. The report noted the strong support for uniform provisions for all corporate and individual insolvencies, at least as regards the liquidation process and subject to appropriate differences for banks, insurance companies and others where such differences were justified by structural requirements,[148] and that uniform provisions for individuals and companies had been developed under the auspices of the Standing

[142] SALC Report, para 3.6.

[143] See, for example, Evans, R G, "Friendly Sequestrations, the Abuse of the Process of Court, and Possible Solutions for Overburdened Debtors" (2001) *SA Merc LJ* 485; Evans, R G, "The Abuse of the Process of the Court in Friendly Sequestration Proceedings in South Africa", 2002 IIR 13.

[144] SALC Report, para 3.6.

[145] *Ibid.*

[146] *Ibid*, para 4.6.

[147] *Ibid*, para 1.4.

[148] *Ibid*, para 5.1. Such differences also exist in Scotland and indeed in many other jurisdictions.

Advisory Committee of Company Law and would be available in the near future.[149] As indicated above, the Standing Advisory Committee of Company Law had remitted this matter to CACIL, which produced its *Final Report Containing Proposals on A Unified Insolvency Act* ("the first CACIL Report") in January 2000. The SALC Report noted that the project committee was of the view that the review of corporate insolvency should be finalised simultaneously with the review of insolvency law and that the SALC itself appreciated the importance of corporate insolvencies and the benefits of uniform legislation.[150]

The first CACIL Report and annexed draft Insolvency and Business Recovery Bill takes as its basis the SALC's draft Insolvency Bill as it stood at December 1999 but clearly identifies proposals which are in conflict with those made by the SALC and proposals original in the context of the SALC's review. The report reflects input from a symposium held in October 1998, a series of workshops held in December 1998 and a conference held in October 1999 as well as consultation with, and written submissions from, interested parties and accordingly states that "it would not be out of place to state that this document reflects the views of the majority of the role players in insolvency law in South Africa."[151] The report recommends the adoption of unified insolvency legislation on the basis that the current legislative arrangements cause problems in practice and that

> [t]here is no sound philosophical reason for divorcing the provisions relating to the winding-up of juristic persons from those relating to individuals. Although there are certain material differences between an individual and juristic persons . . . these differences are not paramount to the extent that the liquidation (of what in both cases amounts to the administration of an insolvent estate) should be governed by separate provisions.[152]

The report notes that South Africa has traditionally been classified as a pro-creditor system since its insolvency law is aimed at protecting creditors rather than assisting struggling debtors,[153] but that the proposed unified legislation "will lean more toward a pro-debtor system than before, but still with the requisite protection for creditors."[154] This is because of the inclusion of provisions designed for the relief of the debtor including provisions on rehabilitation, pre- and post-liquidation compromises and judicial management.[155]

The proposed unified legislation applies to all debtors, including the specialised institutions, such as banks and insurance companies, to which special regimes apply, with appropriate provision to reflect the necessary specialities being provided for within the unified legislation. Similarly, separate provision is made for the different circumstances in which individuals and corporate entities may be liquidated by the court where such differences have been retained.[156] The general process of administration of the liquidated estate is made the same for all debtors, subject to separate provision where necessary to reflect the type of debtor. The concept of voluntary liquidation is retained and the new "liquidation resolution" procedure proposed to initiate it is extended to all those entities capable of making use of it. In a major shift in policy towards rescuing businesses instead of liquidating them, the proposed unified legislation

[149] See *ibid*, paras 5.1 and 5.4.

[150] *Ibid*, paras 5.2 and 5.3.

[151] First CACIL Report, para 1.12.

[152] *Ibid*, para 3.6.

[153] *Ibid*, para 3.8.

[154] *Ibid*, para 3.9.

[155] *Ibid*.

[156] For example, the requirement of advantage to creditors in the liquidation of an individual.

includes business rescue provisions. Since no consensus on the form new business rescue provisions should take could be reached before the report was submitted, the proposed unified legislation simply incorporated a slightly improved version of judicial management, but extended to some other types of entities such as close corporations and trusts,[157] in the meantime.[158] In addition, the provisions of Act 61 of 1973, section 311, were incorporated into the proposed unified legislation. The proposed legislation includes new provisions on the personal liability of directors in the form of a new provision on insolvent trading, but the issue of disqualification and personal liability of directors generally and the possible extension of such provisions to others, such as trustees of a trust or the members of a close corporation, are still under review by the Standing Advisory Committee on Company Law.

The second CACIL Report was produced in response to a request from the Department of Justice and the Law Society of South Africa ("LSSA") to facilitate a research programme with a view to reforming the administration order procedure under Act 32 of 1944 following calls for reform of that procedure from practitioners and other interested parties. The report explains that the procedure has already been amended on a number of previous occasions and was replaced with a new section 74 in 1976.[159] The report includes a proposed Act on Administration, Composition and Related Matters and a proposed Act on the Regulation of Administrators and reflects written submissions previously received by the Department of Justice and input received during and after two conferences/workshops held in September 2000 and January 2001.[160] It adopts a holistic approach to reform on the basis that administration orders form an integral part of insolvency law and can be classified as a debt relief measure[161] and considers the operation and reform of the debt relief measures available to consumer debtors as alternatives to sequestration procedures that might not be available to them.[162]

The report considers two distinct categories of issues: practical issues and issues relating to perceived abuse of the procedure and other matters. It explains that as a result of the variety of sometimes contradictory views put forward on some issues, a number of important issues remain unresolved,[163] but it makes a number of firm recommendations for reform and offers possible alternative options for reform in relation to unresolved issues. In particular, it identifies the following as minimum requirements for reform of the administration procedure itself:[164]

(i) The extension of the period of notice of the application to creditors.

(ii) The possible raising of the current monetary limit of R 50,000. The report suggests, however, that if the limit is raised significantly, consideration would need to be given to applying certain provisions of insolvency law, such as those relating to impeachable transactions and unexecuted contracts, to the administration procedure, which might then be more appropriately included in the new insolvency legislation.[165]

(iii) The functioning of the debtors' courts with particular regard to uniform practices.

(iv) The basis for an application for administration.

[157] First CACIL Report, para 4.18.

[158] See further below.

[159] Second CACIL Report, para 6.1.

[160] *Ibid*, i.

[161] *Ibid*, para 1.15.

[162] *Ibid*, para 1.20.

[163] *Ibid*, p 454.

[164] *Ibid*, p 455.

[165] *Ibid*, p 457. The proper place for the administration order procedure is discussed further below.

(v) The regulation of the office of administrator. The report discusses a number of possible models for a regulatory regime.[166] The proposed Act on the Regulation of Administrators, which is intended to serve as a basis for discussion of the options for regulation of administrators,[167] is based on a scheme of registration of administrators. It provides for the establishment of a Council for Administrators to exercise control over the occupation of administrator. No-one, with the exception of an attorney or the employee of an attorney, is permitted to act as an administrator without first being registered as an administrator and there is provision for a public register of administrators. Specified persons are disqualified from being registered as an administrator and there is also provision for prescribing required qualifications for administrators. There is also provision for a Code of Conduct, penalties for improper conduct including withdrawal or suspension of registration and withdrawal of registration by the Council or the court on grounds other than improper conduct. Administrators will be required to keep separate trust accounts.

(vi) The inclusion of *in futuro* debts within the scope of an administration order and the capitalisation of interest.

(vii) A limited time-period for repayment of the debt.

(viii) The introduction of some form of discharge for the debtor.

(ix) Consequential amendment of other problem areas.

In addition, the report suggests that statutory intervention is also required in related matters such as the regulation of credit and the introduction of a national debt register, the regulation of credit bureaux, debt counselling and debtor education programmes.[168]

The report also considers questions of structure. With regard to the proper place for a revised administration order procedure, it identifies three options: retain it in Act 32 of 1944, include it in the new unified insolvency legislation or enact separate legislation,[169] but it does not make any firm recommendation.[170] With regard to the relationship between a revised administration order procedure and the new composition procedure proposed in the SALC report, it has already been noted that it was recommended that the latter should be adopted subject to clarification of its relationship with the former;[171] the inter-relationship of the various debt relief measures, including sequestration, is seen as very important.[172]

As noted above, reform of the administration procedure is now under consideration by the SALC, which has taken the second CACIL report as a basis for its work, and a discussion paper is expected in the spring of 2005.

The Consumer Credit Bill published for comment by the South African Department of Trade and Industry in August 2004 resulted from a review of credit law carried out by a Technical Committee appointed by the Department of Trade and Industry in March 2002 which considered, *inter alia*, aspects of over-indebtedness[173] and, as noted above, it includes certain provisions on debt re-structuring for consumer debtors. The associated policy framework, Making Credit Markets Work: A Policy Framework for

[166] *Ibid*, p 456.

[167] See Proposed Act on the Regulation of Administrators.

[168] Second CACIL Report, p 455.

[169] *Ibid*, p 455.

[170] The report notes that the answer will depend, at least in part, on who is to supervise the administration: see *ibid*, para 10.1.

[171] See above.

[172] Second CACIL Report, para 10.2.

[173] See the Credit Law Review published by the South African Department of Trade and Industry in August 2003, available at www.thedti.gov.za. This in turn drew on several earlier reports.

Consumer Credit ("the policy framework"),[174] states baldly that "South African law provides no effective protection against over-indebtedness and insufficient rehabilitation mechanisms to assist people who have become over-indebted".[175] To address this, the policy framework first outlined a number of new provisions designed to prevent and penalise reckless lending, including a requirement on courts to obtain a report from a debt counsellor before making a final decision in respect of, *inter alia*, an administration order,[176] but went on to say that beyond such measures "there is still a need to make provision for some form of relief, other than more extreme measures such as debt administration, for those who are unable to repay their debts".[177] It envisaged such provision taking the form of the creation of a national network of regulated debt counsellors who would be able to renegotiate a consumer's debt commitments with a view to providing relief in cases of over-indebtedness, with appropriate mechanisms for debt re-scheduling and debt reduction which would be applied in accordance with specified criteria and a prescribed approach being built into the scheme[178] and this is reflected in the relevant provisions of the Consumer Credit Bill. The key features of these provisions are:

(i) An application by a consumer to a debt counsellor to rearrange obligations under one or more credit agreements or to have one or more credit agreements declared to be reckless credit.

(ii) An assessment by the debt counsellor of the consumer's financial ability to meet his or her obligations in a timely manner and of whether any credit agreement constitutes reckless credit.

(iii) A discretion for the debt counsellor to make appropriate recommendations to the new Consumer Tribunal to be established under the provisions of the bill where the debt counsellor concludes the consumer's obligations cannot be met in a timely manner. Possible recommendations are a recommendation that one or more credit agreements be declared to be reckless credit or a recommendation that one or more of the consumer's obligations be restructured by extending the period of the agreement and reducing each payment accordingly, postponing during a specified period the dates on which payments are due, extending the period of the agreement and postponing during a specified period the dates on which payments are due and/or recalculating the obligations because of contraventions of certain other specified provisions.

(iv) A hearing on any such recommendations by the Consumer Tribunal, which may decide to reject the recommendations or to make an order declaring any credit agreement to be reckless and suspending it or an order restructuring the consumer's obligations in any of the ways previously set out.

(v) Restrictions on the consumer incurring new credit pending a decision on an application *or* during the period of the approved rearrangement.

(vi) Restrictions on enforcement by credit providers pending a decision on an application *or* while an approved rearrangement is being implemented.

There are obvious similarities between these proposed provisions and particular features of the new DAS in Scotland;

The policy framework goes on to acknowledge the link with insolvency law, stating:

[174] Available at www.thedti.gov.za.

[175] Policy Framework, preamble to Chapter Six, Dealing with Debt.

[176] *Ibid*, para 6.5.

[177] *Ibid*, para 6.10.

[178] *Ibid*, para 6.11.

> [t]he Credit Law Review highlighted serious weaknesses in the insolvency legislation and particularly in terms of the mechanisms for personal insolvency of consumers who are over-indebted. This is consistent with the findings of the Borainne Commission, which was established by the Department of Justice.[179]

It continues

> Government recognises that there are weaknesses in the insolvency law and that measures must be implemented to rectify these weaknesses in order to create effective mechanisms for personal insolvency and debt rehabilitation. This forms part of a holistic strategy for dealing with the problems resulting from over-indebtedness.[180]

This reflects the approach of the second CACIL report, with its emphasis on a holistic approach.

## COMMON THREADS IN REFORM

It is clear that there are a number of common threads running through the recent and further proposed reforms in Scotland on the one hand and the proposed reforms in South Africa on the other. There are also some interesting differences in approach to some of these issues which merit exploration.

One common thread is business rescue. Company insolvency law reform in the UK has for some time placed an increasing emphasis on rescue rather than liquidation – the White Paper specifically states in the introduction to the company insolvency law proposals that:

> For the last twenty-five years or so, the focus of insolvency law reform in the United Kingdom has increasingly been on the promotion of a rescue culture, a trend which started with the work of the Cork Committee . . .[181]

A move towards business rescue is now also evident in the proposed reforms in South Africa. The first CACIL report describes the inclusion of compositions, compromises and arrangements and other business rescue provisions in the proposed unified insolvency legislation as "one of the main thrusts of the new proposals, and . . . a definite shift towards saving businesses instead of liquidating them".[182] The report states that South Africa does not (yet) have a culture of business rescue, but that it "has started to realise the need for effective business rescue provisions under new insolvency legislation".[183]

Such a shift in emphasis is not, of course, problem-free. The report notes that:

> One of the problems which will be encountered in introducing business rescue provisions in South Africa, is that the South African economy is vastly different from the countries which presently have effective business rescue regimes. For this reason, it is doubtful whether another country's business rescue provisions can merely be imported into our legal system. If one is to introduce a business rescue regime in South Africa, such provisions would have to be tailor-made for the local circumstances.[184]

It continues:

[179] *Ibid*, para 6.14. The findings of the Borainne Commission are the second CACIL Report.
[180] *Ibid*, para 6.15.
[181] White Paper, para 2.1.
[182] First CACIL Report, para 4.18.
[183] *Ibid*, para 4.20.
[184] *Ibid*, para 4.21.

> Another problem is getting major creditors, such as banks, to catch on to the idea of business rescue. More often than not, a business rescue regime means that some creditors somewhere are going to lose out on the immediate payment of their claims, and it is going to take a concerted effort to get these creditors to catch on to the idea. No business rescue regime can hope to be successful in South Africa without the active participation of big business.[185]

It might also be added that a cultural shift in the attitude of the insolvency and related professions would also be required and that it would inevitably take time for the relevant expertise to develop. Furthermore, concern has been expressed that other reforms, in particular labour law reforms to strengthen the rights of employees, might impact unfavourably on the development of a rescue culture in South Africa in practice.[186] These problems should not, however, be insuperable.

It was noted above that no consensus on the form that any new business rescue provisions should take could be reached before the first CACIL report was submitted and the proposed unified legislation therefore incorporated only a slightly improved version of judicial management meantime. The report went on to propose, however, that work on new business rescue provisions should continue with a view to their later inclusion in either the unified insolvency legislation or a separate Act.[187] Such work has in fact been ongoing and it is understood that proposals for a new business rescue regime will be submitted together with the report of the Commission of Enquiry on the Liquidations Industry referred to above. A more rescue-oriented approach therefore seems to have been accepted in principle, although it is clear that there are still practical hurdles to be overcome.

It is notable that the move towards business rescue is not confined to company debtors in South Africa. As discussed above, the improved judicial management provisions incorporated into the proposed unified insolvency legislation are made applicable to some other types of entities such as close corporations and trusts and the first CACIL report notes that "[a] major new approach is that the proposals relating to business rescue are not only aimed at companies, but at other business entities such as close corporations, trusts, partnerships and, as some have suggested, also to individuals".[188] It is suggested that such an approach would merit consideration in Scotland. At present, the statutory business rescue procedures in the IA 1986, *ie*, CVAs and administration, are available only to companies and LLPs in Scotland and not to other types of business entity or to individuals in business.[189] Similarly, section 425

[185] *Ibid*, para 4.22.

[186] See, for example, Boraine, A, Van Eck, BPS and Burdette, DA "The amendments to sections 38 and 98A of the Insolvency Act, and section 197 of the Labour Relations Act", 2002 SAILR 171 and the literature there cited. Similar issues have arisen at the interface of insolvency law and employment law in the UK.

[187] First CACIL Report, para 4.26.

[188] *Ibid.*

[189] *Cf* the position in England and Wales, where a modified version of the CVA and administration procedures under the IA 1986 are open to partnerships other than LLPs under the Insolvent Partnerships Order 1994, SI 1994/2421; the changes to the administration procedure made by the EA 2002 do not currently extend to partnerships other than LLPs in England and Wales, but the government is considering whether so to extend them. Other business debtors in England and Wales, however, are at present in a similar position to their counterparts in Scotland in terms of the availability of business rescue procedures as such and the EA 2002 reforms to bankruptcy law in England and Wales, although aiming to encourage entrepreneurship, concentrated on reforming the formal process of bankruptcy itself rather than providing alternative business rescue procedures which might prevent bankruptcy in the first place. The Department of Trade and Industry's White Paper *Fair, Clear and Competitive: The Consumer Credit Market in the 21st Century*, Cm 6040, published in December 2003, stated that the Insolvency Service and the Department of Constitutional Affairs were working together to consider whether the current statutory systems available to debtors were sufficient and in particular whether there was a need for a non-court-based resolution to debt problems for those who were unable, for whatever reason, to go through existing channels and that there would be consultation on viable solutions early in 2004 (para 5.79). A consultation paper entitled *A Choice of Paths: Better options to manage over-indebtedness and multiple debt* ("A Choice of Paths") was duly published in July 2004, but the proposals, in respect of some of which a further consultation is expected shortly, seem to be designed primarily with consumer rather than business debtors in mind. It is understood,

arrangements are available only to companies and LLPs. The DAS is available to natural persons, including those in business, but not to other types of debtor and in any event it is not designed as a business rescue procedure, although it is possible that concluding a DPP might allow an individual to continue in business where otherwise he or she could not.[190] Compositions at common law and trust deeds for creditors can also be entered into with a view to effecting a business rescue, but again these procedures are not designed for that purpose and in practice are difficult to utilise in that way. Yet the legal form of a business may be fortuitous and there is therefore a strong argument that appropriate rescue procedures should be available to a business whatever its form. As noted above, the First Consultation Paper sought views on whether the existing and proposed debt solutions would provide appropriate solutions for all debtors, but although the lack of specific rescue procedures for trading debtors other than companies and LLPs was raised by some consultees, this has not been taken up in the Second Consultation Paper.[191] It is suggested, however, that Scotland should give serious consideration to following the proposed South African approach in this matter and making suitable business rescue procedures available to all types of business debtor.[192]

A second common thread is a reduction in the number of categories of preferential creditor. This has now been achieved to a significant extent in the UK with the abolition of Crown preference effected by the EA 2002, the principal remaining category of preferred/preferential creditor now being employees for certain limited sums. The White Paper presented this reform as an integral part of the package of reforms ultimately given effect to by the EA 2002, describing it as "a step which will bring major benefits to trade and other unsecured creditors, including small businesses".[193] It is therefore linked with business rescue. The White Paper noted simply that:

> Preferential claims originated in the late 19th century, but in recent years the trend in other jurisdictions been towards restricting or abolishing Crown or State preference as, for instance, in Germany and Australia. We believe that this is more equitable.[194]

This reform built on earlier reforms. In Scotland, the SLC in its report in 1982 had identified as one of the objectives behind its recommendations for reform the promotion of equality among the creditors as a class by, *inter alia*, reducing the categories of preferential creditors. It recommended the abolition of the then Crown preferences for income tax, capital gains tax, corporation tax, PAYE,[195] VAT,[196] car tax, betting and gaming duties, development land tax and social security contributions and the then local authority preference for local rates and further recommended that the proposed abolition should extend to both bankruptcy and liquidation. The *Cork Report* had also identified one of its main objects of reform as being to ensure a fairer

however, that changes to the provisions relating to IVAs in England and Wales which would result in different versions of the IVA procedure specifically tailored for consumer and business debtors are under consideration although no formal proposals in this respect have been made. As noted above, however, IVAs are in any event available only in England and Wales and not Scotland.

[190] *Cf* the statutory IVA procedure in England and Wales, which is open to individuals whether in business or not, but which in practice does not operate very effectively as a rescue procedure for business debtors.

[191] See above.

[192] The related issue of whether there should in general be different regimes for individual business and consumer debtors is discussed further below.

[193] Foreword to the White Paper.

[194] White Paper, para 2.19.

[195] Pay As You Earn, the income tax deducted by an employer from an employee's wages or salary for onward transmission to the taxing authorities.

[196] Value Added Tax.

distribution of the assets realised in the course of the insolvency proceedings and so to allay existing dissatisfaction in this area by, *inter alia*, a substantial reduction in the number and extent of preferential claims. It took a slightly more conservative approach, however, recommending abolition of the preferences for income tax, corporation tax, capital gains tax, development land tax and local rates but reduction rather than abolition of the preferences for PAYE, VAT, car tax, betting and gaming duties and social security payments. Despite government resistance, the B(S)A 1985 and the IA 1986 reduced the list of preferred/preferential creditors, but a significant number of Crown preferences still remained to be abolished by the EA 2002.

The same desire to reduce the number of categories of preferential creditor for reasons of equity is evident in the proposed reforms in South Africa. As noted above, the SALC has proposed the abolition of most of the preferences currently provided for in Act 24 of 1936, most notably those in favour of state institutions, and the retention of only those preferences payable out of the free revenue of the estate in favour of employees, contributions to employee funds and claims for arrear maintenance payable in terms of a court order. The SALC had previously made similar recommendations in its report on the review of preferent claims in insolvency in 1984, in which it took the view that it should be accepted as a general premise that unsecured creditors should be dealt with on an equal footing except for justified preferences; justified preferences should be based on considerations of fairness; and the mere fact that the State is a creditor does not of itself justify a preference. These recommendations were specifically rejected by the cabinet in 1989, apparently on the basis that the abolition of the preferences for taxes would have unacceptable cost implications for the State,[197] but in its *Working Paper 61 on the Review of the Law of Insolvency*: *Statutory provisions that benefit creditors*,[198] published in 1995, the SALC nevertheless maintained its position that the mere fact that the State is a creditor does not of itself justify a preference, arguing that the possible loss of state revenue does not outweigh the unfair effect which existing preferences have on ordinary creditors and that there is no logical reason why other creditors should have to bear the losses in state revenue. The SALC report notes that most commentators on that working paper supported the view that statutory provisions preferring creditors are undesirable and cannot be justified merely because revenue is utilised for the benefit of the public or because the state or state-assisted bodies are involved,[199] and concludes that "the case for abolition of the incongruous and over-long list of preferent rankings in the Insolvency Act has not been altered".[200]

A third common thread is the provision of effective alternatives to formal insolvency procedures for individual debtors. In Scotland, this can be seen most obviously in the introduction of the DAS, which forms an important element in the Scottish Executive's new approach to debt management and enforcement[201] and which was based on a perceived need for a simple mechanism for dealing with "the increasing social problem of multiple debt".[202] The DAS was not, of course, a new idea: as noted above, the SLC had proposed the introduction of a debt arrangement scheme as far back as 1980.[203] It had, even then, identified what it called a gap in the provision made by Scots law

[197] SALC Report, para 80.3.

[198] The Working Paper included consideration of statutory provisions in favour of creditors outwith Act 24 of 1936 as well as the preferent claims provided for by that Act itself.

[199] SALC Report, para 80.10.

[200] *Ibid*, para 80.11.

[201] Policy Memorandum on the Debt Arrangement and Attachment (Scotland) Bill, SP Bill 52-PM, paras 12, 15.

[202] *Ibid*, para 15.

[203] Scottish Law Commission *Fourth Memorandum on Diligence: Debt Arrangement Schemes*, Scot. Law Com. Consultative Memorandum No 50 (1980).

for helping wage earners in multiple debt to arrange for payment of those debts to the benefit of both debtor and creditors[204] and provisionally concluded that it would be desirable to introduce "a relatively simple and inexpensive process" which would allow such debtors to make orderly and regular payments of their debts to their various creditors.[205] It had recommended the introduction of such a scheme in 1985[206] although it was not implemented at the time.[207] It was also noted above that the First Consultation Paper had emphasised the need to provide a comprehensive debt management framework with appropriate solutions for all debtors[208] and that the Second Consultation Paper has acknowledged that there are a number of debtors with little or no assets or income for whom not only bankruptcy but the other existing solutions may be inaccessible or inappropriate,[209] a situation which led the Scottish Executive to establish the Working Group on Debt Relief referred to above to examine, *inter alia*, how extensive the problem is and what possible solutions might be.[210]

A similar move towards providing effective alternatives to formal insolvency procedures for individual debtors in South Africa can be seen in the SALC's proposal for a pre-liquidation composition, the recommendations of the second CACIL report and the provisions for debt re-structuring included in the Consumer Credit Bill, and for similar reasons. The SALC report states simply that "[p]rovision must be made for debtors with little or no assets who through no fault of their own are unable to pay their debts", pointing out that "[i]f a debtor obtains money somewhere or is able to conclude satisfactory arrangements for the payment of debts, a composition may be more advantageous for both the debtor and creditors than the liquidation of the debtor's estate".[211] The second CACIL report highlights the expanding use of credit in South Africa[212] and the consequence that "many consumer debtors end up in a debt trap".[213] It notes that for many "the luxury of the rather expensive sequestration procedures is simply not available or applicable due to a lack of realisable assets. It is also questionable whether sequestration is the answer in many cases of default".[214] It concludes that "simple, rather inexpensive procedure(s) is/are needed to deal with many

[204] *Ibid*, para 1.2.

[205] *Ibid*, para 1.4.

[206] Scottish Law Commission *Report on Diligence and Debtor Protection*, Scot. Law Com. No 95 (1985). The report contained detailed provisions for the scheme, some, but not all, of which have been adopted in the new DAS.

[207] It is notable that the Cork Committee had also recommended the introduction in England and Wales of debt arrangement orders, designed to provide a simple, accessible and inexpensive procedure for dealing with the ordinary consumer debtor whose affairs did not require investigation and who had no significant realisable assets but did have the prospect of being able to meet part or all of his debts out of future income, but its specific proposals in this respect were not accepted either, although changes to the law relating to voluntary arrangements between a debtor and his creditors designed to facilitate such arrangements and make them a more effective alternative to bankruptcy were made in the form of the provisions on IVAs now contained in Part VIII of the IA 1986: see A Revised Framework for Insolvency Law, 1984, Cmnd 9175, paras 136–139. As noted above, the recent consultation paper *A Choice of Paths* put forward a number of proposals designed to help deal with individuals who are over-indebted and, in particular, those in multiple debt in England and Wales: see note 189 *supra*. These proposals included the introduction of a *No Income No Assets* ("NINA") debt relief scheme to be administered by the Insolvency Service; the introduction of an Enforcement Restriction Order to provide temporary relief for debtors who are unable to meet their commitments in the short to medium term; reform of the existing court-based administration order scheme; and the strengthening of private and voluntary sector repayment schemes, including the possible introduction of provisions for binding dissenting minority creditors and composition. The responses to the consultation have not been published at the time of writing, but a further consultation on the NINA debt relief scheme is expected to be issued shortly.

[208] First Consultation Paper, para 6.24.

[209] Second Consultation Paper, para 7.8.

[210] *Ibid*, para 7.10.

[211] SALC Report, para 124.1.

[212] "Within the modern South African society the micro-lending business as well as credit cards and related forms of credit is expanding": Second CACIL Report, para 4.1.

[213] *Ibid*, para 4.1.

[214] *Ibid*, para 4.2.

of these cases"[215] and specifically draws a parallel in this respect with similar sentiments expressed by the Cork Committee.[216] As noted above, the policy framework for the Consumer Credit Bill also acknowledges that South Africa has no effective protection against over-indebtedness and insufficient rehabilitation mechanisms to assist people who have become over-indebted and the corresponding need to make provision for some form of relief for those who are unable to repay their debts other than the existing formal procedures such as administration orders and sequestration.

In both jurisdictions, therefore, the move towards providing effective alternatives to formal insolvency procedures for individual debtors is driven by the growing problem of consumer debt arising from the now widespread availability of credit, although the underlying causes of that debt problem differ in some respects in the two jurisdictions.[217] In South Africa, the second CACIL report notes that one of the main issues facing South African law is that it has to deal with a sophisticated first world economy and a third world economy.[218] The report goes on to explain that:

> Within the ambit of consumer debt, it must be remembered that many ordinary citizens rely on credit in order to obtain consumer goods on the one hand but many, especially those from the previously disadvantaged communities, have a need for credit in order to pay funeral expenses of family members or even just to be able to maintain themselves.[219]

In Scotland, it is believed that debt problems arising for the second of these reasons are much rarer than in South Africa; although not unknown.[220] It is arguable, however, that the underlying cause of the debt problem should make no difference to the general availability of appropriate debt relief measures, although it may be relevant to particular aspects thereof, such as the availability of a discharge.[221] Addressing the underlying causes of debt is, of course, a separate issue.

A fourth common thread, of which the provision of effective alternatives to formal insolvency procedures for individual debtors may be regarded as one particular strand, is the provision of an opportunity for a fresh start for the debtor. However, it is possible to discern a difference in approach to this issue in the two jurisdictions, which is worthy of exploration. In South Africa, a debtor may ultimately obtain a fresh start through the sequestration process but there is at present no generally effective method of doing so outwith that process. The SALC report seeks to provide an opportunity to obtain a fresh start outwith the sequestration process through its proposals for a pre-liquidation composition, which is designed to provide the opportunity for a fresh start to debtors for whom it is not available through the sequestration process because access to the sequestration process is effectively denied to them as a result of the requirements for sequestration. However, it intentionally makes no change to the requirements of sequestration itself. Instead, it seeks to achieve a situation where a debtor who can access the sequestration procedure can obtain a fresh start through that mechanism and a debtor who cannot (or does not wish to) access the sequestration

[215] *Ibid*, para 4.2.

[216] See note 207 *supra*.

[217] This trend is one which is also evident in other jurisdictions: see Ziegel, J *Comparative Consumer Insolvency Regimes* (Oxford: Hart Publishing, 2003) and Niemi-Kiesilainen, Ramsay & Whitford (eds), *Consumer Bankruptcy in Global Perspective* (Oxford: Hart Publishing, 2003).

[218] Second CACIL Report, para 1.15.

[219] *Ibid*, para 4.2.

[220] In the debates surrounding the APWSA 2000, evidence was given to the committees scrutinising the Bill to the effect that in the poorest sections of Scottish society, there were cases where credit was incurred for necessities rather than luxury consumer goods and a recent report by Citizens Advice Scotland, *On the Cards: The debt crisis facing Scottish CAB clients* (January 2004) suggests that people on low incomes may use credit "in their financial struggles to get by": para 66.

[221] This is discussed further below.

procedure can nonetheless obtain a fresh start by availing himself of an effective alternative thereto. The second CACIL report does not accept that the SALC proposals would in fact achieve this in their current form,[222] but adopts a similar approach in principle in so far as it recommends the provision of effective alternative procedures to formal insolvency which will provide a fresh start to those debtors to whom such is not available through the formal insolvency procedure. The same approach is evident in the policy framework for the Consumer Credit Bill with its debt re-structuring provisions. The emphasis is clearly on the provision of a fresh start to debtors through effective alternatives to formal insolvency procedures while leaving the sequestration process itself at least unchanged.

In Scotland, a debtor may obtain a fresh start through the sequestration process[223] and there are also specific mechanisms for obtaining a fresh start outwith sequestration in the form of trust deeds for creditors and, now, the DAS.[224] As noted above, the introduction of the DAS encapsulates the same basic approach to the issue of a fresh start for the debtor as is evident in the South African proposals, that is, the provision of effective alternatives to sequestration. This approach is also discernible in the First Consultation Paper's emphasis on the need for a comprehensive debt management framework with appropriate solutions for all debtors[225] and in the continuing search for possible solutions to the problem of those debtors for whom not only bankruptcy but other existing solutions may be inaccessible or inappropriate.[226] In contrast to the South African proposals, however, in Scotland it is also intended to make radical changes to sequestration itself for the purpose of facilitating a fresh start for (non-culpable) debtors. Furthermore, these changes are intended to extend to consumer debtors, notwithstanding that they were conceived primarily for the purpose of encouraging entrepreneurship by making it easier for those who have failed in business honestly to try again[227] and that consumer debtors might be regarded as raising different policy issues.[228] At the same time, it is intended to make access to sequestration by debtors easier in some respects and to refrain from imposing any conditions requiring a debtor to divert to a suitable alternative procedure such as the DAS. The Second Consultation Paper makes it clear that the changes to sequestration are not intended to make it an easy option for debtors,[229] but it may nonetheless

[222] Second CACIL Report, para 5.52.

[223] And generally speaking may do so more easily than in South Africa, both in terms of accessing the sequestration procedure itself and in terms of the basis on which a discharge is available. As noted above, in Scotland it is also more difficult for a debtor to access the sequestration procedure than for a creditor to do so, but there is no requirement of advantage to creditors, the biggest hurdle to access to sequestration in South Africa, and discharge is available automatically after 3 years unless deferred on good cause.

[224] Although both mechanisms may be regarded as having certain limitations in terms of providing a debtor with a fresh start: a trust deed does not automatically provide a debtor with a discharge, although in practice it will normally do so, usually in similar terms to the discharge provisions in sequestration *ie*, discharge after three years irrespective of how much, if anything, has been paid to the creditors; and the DAS does not have any maximum time limit and does not provide for the composition of debts unless individual creditors agree.

[225] First Consultation Paper, para 6.24.

[226] Second Consultation Paper, para 7.8, discussed above.

[227] See above.

[228] This was an issue which proved very controversial in England and Wales, where many respondents to the White Paper were concerned about the extension of the reforms to consumer debtors on the basis that the increased risk-taking which they were designed to encourage was undesirable in consumers and where repeated (unsuccessful) attempts were made to amend the EA 2002 during its passage through parliament to provide for a distinction to be made between business and consumer debtors, particularly with regard to the minimum period to elapse before discharge. The issue does not seem to have generated quite the same degree of controversy in Scotland where, despite the fact that the introduction of a one-year period bankruptcy period was the issue on which consultees were most divided, this particular aspect did not attract the same level of comment.

[229] Second Consultation Paper, para 5.10.

appear to them to be a more attractive option than the alternatives,[230] and combined with easier debtor access to sequestration and an unfettered choice of process, the result may well be that more debtors will be able to, and will prefer to, opt for sequestration despite the Scottish Executive's wish to encourage the use of the alternative procedures it has been at such pains to provide.

A fifth and final common thread, and perhaps the most critical of all, is the need to take an integrated approach to reform and to articulate clearly the inter-relationship of the various procedures. In Scotland, as noted above, one of the drivers for change identified in the First Consultation Paper was the importance of having an integrated debt management framework within which the debt management tools available would work together to form a comprehensive package of solutions and the reform proposals specifically address the question of integration, if to a limited extent.[231] In South Africa, as noted above, the second CACIL report emphasised the importance of the relationship between administration orders, the SALC's proposed pre-liquidation composition and sequestration; indeed, it noted that there was "a question whether the proposed pre-liquidation composition cannot replace the administration order in its current form since exactly the same results can be effected by such a composition whilst it opens the door to other forms of relief as well".[232] In Scotland, a similar issue might be whether there is a continuing need for trust deeds for creditors following the introduction of the DAS.[233] Whatever the procedures, however, the critical issue is that their inter-relationship is clear.

## CONCLUSION

Drawing on the common threads identified above, and taking an overview of insolvency law reform as a whole, it is suggested that it is possible to identify a number of basic elements that might form a common framework for insolvency law reform in both jurisdictions, although of course the precise form of such elements is likely to differ in each jurisdiction in order to take account of the differing legal and economic contexts. These elements are:

(i) provision of appropriate business rescue procedures for all types of business debtor, including individuals in business;

(ii) reduction of the number of categories of preferential creditors;

(iii) provision of effective alternatives to sequestration for individual debtors, including individuals in business where it is not possible or desired to rescue the business but an alternative procedure would be more appropriate than a liquidation-type procedure;

(iv) retention of a procedure for liquidating the assets of a debtor in appropriate cases, including appropriate provision for a fresh start for individual debtors undertaking this procedure;

[230] For example, as noted above, it is proposed that debtor contributions from income will be limited to a maximum of three years in a sequestration, whereas there is no maximum time limit on payments under the DAS (although one may be agreed as part of the DPP). There are, of course, other considerations, such as the fact that sequestration will still impose certain restrictions on the debtor, but as it is also proposed that these will be considerably reduced, the debtor may not be particularly concerned about these.

[231] As discussed above, the proposals have concentrated primarily on the question of the integration of sequestration and the new DAS, but little attention has been paid to the other alternatives such as trust deeds and how they are to fit into an integrated system.

[232] Second CACIL report, para 10.2.

[233] As discussed above, the reform proposals clearly envisage the retention of trust deeds, albeit subject to some reform, and it is suggested that this is the correct approach since they may be appropriate in different circumstances. The matter is, however, being considered further by the Working Group on Debt Relief referred to above.

(v) an integrated debt management framework with clear delineation of the inter-relationship of the available procedures.

The Scottish proposals for reform incorporate a number of these elements, or part of them, but do not fully address a number of important issues such as the provision of specific business rescue procedures for debtors other than companies/LLPs and proper integration of the available debt management tools, in particular the articulation of sequestration and the alternatives thereto. It is suggested that the reform process ongoing in Scotland should address these matters and that due consideration of the South African proposals in these respects would prove useful. In South Africa, the various proposals for reform also incorporate many of these elements, but to date none have been implemented, although it now appears as if at least some of the proposed reforms are to be taken forward in the near future and it is hoped that the remaining reforms incorporating these elements will also follow in due course.

NOTTINGHAM LAW JOURNAL

VOL 14(1) 2005

# CASE AND COMMENT

*The address for the submission of material for this section is given at the beginning of this issue*

## REFORM OF HOUSING LAW

*Renting Homes (Law Com No 284) (Law Commission, 2003)*

This Law Commission narrative report recommends a major reform of housing law. The report follows an unusually wide and thorough consultation process, including two consultation papers.[1] The proposed reforms are not primarily concerned with the substance of housing law (security of tenure, rent control, right to buy, *etc*); rather the report makes radical recommendations on the form of the law. These recommendations are in turn based on a radical approach to the identification of, and deployment of, appropriate principles for structuring housing law. If implemented the reforms could mark a new era for housing law, an area of law that has been notorious for complexity, obscurity, and confusion.

The report has three guiding policy imperatives. Simplicity: that simplicity of the law should be an important aim, as unnecessary complexity leads to both inefficiency and injustice. Functionality: that housing law should be shaped by the needs of the housing market rather than by constraints of legal form, as the social and economic aspects of housing law are of great importance. Protective: that the law should be avowedly directed towards the protection of tenants interests, as tenants are small-scale buyers (consumers) in a market marked by chronic shortages in supply and inelastic demand.

These policy imperatives give rise to four formal principles. First, that housing law should not be constrained by the institutions of property law (simplicity and functionality). Second, that there should be as few types of tenancy as possible (simplicity and functionality). Third, that the substance of housing law should be ascertainable from the documents issued by landlords to tenants (functionality and protective). Fourth, that housing law should provide the maximum of flexibility to people in the rented housing sector as is compatible with the protection of tenants (functionality and protective).

There are obvious links between these policies and principles. For example, the law can become more responsive to legislative policy, and simpler to ascertain, if complexities emanating from land law are excluded. Such a simplifying exclusion in turn makes conceivable the aim of embedding the substantive law in documents issued by landlords.

The reforms, if successful, will transform housing law. They will detach housing law from land law, and enable it to become a largely self-sufficient area of law. Housing law began as a series of legislative modifications to land law. The proposed reforms

[1] Renting Homes: 1. Status and Security (Law Com 162) (Law Commission, 2002) and Renting Homes: 2. Co-occupation, Transfer and Succession (Law Com 168) (Law Commission, 2002).

start from the position that this historical reliance upon land law is pernicious and must be excised from housing law. When the report seeks principles to guide the reform of housing law it turns to contract law and consumer law.

The report proposes that there should be two types of "occupation agreement", which it calls "type I'" and "type II". Type I will be referred to as "social agreements", and type II will be referred to as "market agreements". Social agreements are modelled on the present secure tenancy governed by the Housing Act 1985, and market agreements are modelled on the assured shorthold tenancy governed by the Housing Act 1988. These two types of agreement will operate in all situations in which premises have been, or are in future, made available, for a valuable consideration, as a home for an occupier for less than 21 years, subject to exceptions. The exceptions are generally familiar, and include service agreements, lodging agreements, and agreements subject to other legislative regimes such as business tenancies and mobile homes.

Social agreements will be for an indefinite period, provide significant security of tenure, and will give limited succession rights to occupiers. Market agreements will be either for a fixed period, or for an indefinite period. Indefinite market agreements will be inherently insecure, being subject to termination by service of a notice to quit by the landlord, unless the landlord amends the agreement to exclude its right to serve notice to quit.

For both types of occupation agreement possession by the landlord will only be possible after service of a notice on the occupier and successful subsequent pursuance of court proceedings. For social agreements the court will always have a discretion to exercise before granting possession. For market agreements there will be two mandatory grounds for possession, rent arrears and service by the landlord of a notice to quit.

In social agreements there will be two important types of ground for possession, occupier misconduct and management reasons. Possession for occupier misconduct will be available for breaches of the agreement, and for more general misconduct including anti-social behaviour in the neighbourhood or domestic violence. It will not be necessary to offer alternative housing to occupiers evicted because of misconduct. Possession for management reasons will include some instances of under occupation, the need to use specialised housing efficiently, and redevelopment schemes. It will be necessary to offer suitable alternative housing to occupiers evicted for management reasons.

In market agreements, possession will be available for occupier misconduct. This will include general misconduct as well as breach of the agreement. Possession will also be available for management reasons, in the same circumstances, and subject to the same limitations as for social agreements. In practical terms, the notice ground for possession will mean that alternative grounds for possession are important only in the context of fixed term market agreements.

Both types of agreement will be composed of clauses of different status, referred to in the report as "terms". Terms may be: key, special, compulsory minimum, default, substitute, or additional. The status of a term is important as it will determine whether a term can be varied by the parties (and if so, how it may be varied), whether a term will be provided in the model agreements that the government will make available, and whether a term will be subject to a test of fairness under the Unfair Terms in Consumer Contract Regulations 1999 ("UTCCR").

Key terms are the identity of the parties, the premises, the time period the agreement runs for (start date and either fixed term or indefinite), and the rent (consideration). These terms will be peculiar to the agreement, and therefore not provided for by model

agreements, and will be the sole terms exempt in principle from the fairness test laid down by UTCCR (core terms). Variation of the different key terms will be subject to different rules set out in compulsory minimum and default terms.

Special terms are imposed on the parties. They are not variable by the parties. They are exempt from UTCCR. Special terms are terms imposed by the state into all agreements for reasons of policy. A proposed example would be a term prohibiting anti-social behaviour by the occupier (including domestic violence) imposed in all agreements (social and market).

Compulsory minimum terms are variable by the parties, but any variation must be in favour of the occupier. They will be provided in the model agreements, and adoption of the model term will automatically satisfy UTCCR. Any alteration to the model terms will be subject to the fairness and transparency tests of UTCCR. Compulsory minimum terms will provide, *inter alia*, for: the grounds for possession proceedings (including a general ground for breach of the agreement by the occupier); the necessity to serve notice before taking possession proceedings; the ability of an occupier under an agreement for an indefinite period to serve a notice to quit; and the occupier's right to request the addition of another party to the agreement, subject to the landlord's consent (not to be unreasonably refused).

Default terms are variable by the parties, and the variation may be in favour of either party. Default terms will be provided in the model agreements, and adoption of the model term will automatically satisfy UTCCR. If a default term is modified the modified term is known as a substitute term. Substitute terms are subject to both the transparency and fairness tests of UTCCR. Default terms will provide, *inter alia*, for: a list of obligations binding on occupiers (breach of any of which will justify possession proceedings); the maximum period for which a notice to quit served by a landlord is effective; the minimum period of notice to be given in a tenant's notice to quit; and the occupier's right to determine who else lives in the premises, not being a party to the agreement.

Additional terms are any other terms agreed by the parties concerning matters not dealt with by key, special, compulsory minimum, or default terms. Obviously they will not be dealt with in the model agreements. All additional terms will have to satisfy both the transparency and fairness tests of UTCCR.

This structure of terms of different status is designed to facilitate the production in occupation agreements of the applicable substantive and procedural law of housing in the text of the agreement itself. The agreement will usually be in model form, based on the lengthy model agreements produced by government. The model form should be clear and easy to use. The occupier will have in a single document an account of the rights and obligations that follow from the agreement, and advisers will be able to advise from a position of experience of the model terms in use. Freedom to negotiate is preserved by the possibility of varying compulsory minimum and default terms, as well as the use of additional terms. Any non-standard term negotiated by the parties (in practice imposed by the landlord) will be subject to the UTCCR, which should encourage fairness and transparency.

Although the report is more important for the effects it would have upon the form of housing law than for its recommendations on substantive questions, its substantive recommendations are important in their own right. There are important recommendations in connection with abandoned premises, tolerated trespassers, succession rights, domestic violence and anti-social behaviour, the availability of mandatory grounds for possession, structuring judicial discretion, and supported housing schemes. Particularly welcome to the author are the recommendations for reform of the law governing joint

tenants of rented accommodation.[2] Generally, the report is a model of excellence in the way it deals with these issues. However, it is in terms of the proposed new structure for the law that the report must be judged, it is in this respect that it is both most ambitious, and potentially most important for the future of the law.

A disturbing feature of the report is the effect upon property rights of personal misconduct, given the recommendations relating to anti-social behaviour. A proposed special term will make conduct that would not otherwise be relevant to an occupation agreement a ground for possession. The policy behind the special term does not originate with the Law Commission. Powers provided to allow landlords to control the anti-social behaviour of tenants provoked most controversy in the consultation process. Fundamentally, the punishment of anti-social behaviour by means of the deprivation of property rights, without the protection of criminal procedures, is a questionable policy.

Concern seems appropriate over the costs placed upon landlords in producing and updating the occupation agreements. In particular the recommended power of the Secretary of State to amend compulsory minimum terms by statutory instrument is likely to place a not inconsiderable burden upon landlords of monitoring the law for relevant changes.

The assumption that it is a coherent strategy for reform to ignore land law also deserves closer consideration. Such a strategy accepts the existence of parallel legal devices in respect of a single agreement, *ie* occupation agreements and leases or licences. The existence of such parallel devices will tend to generate exactly the type of interference problem the Law Commission tried to avoid by turning to contract and consumer law. The report hardly touches upon the effects upon occupation agreements of a transfer of the landlord's interest in the land (see paras 8·92(2) and 8·66). It is assumed that the pre-existing common-law of leases and licences will function much as it does currently. Yet the proposed duration of all social agreements, and many market agreements, will be for an indefinite period. This could lead to a transformation of the application of the common-law in this area, with licences ousting leases as the predominant common-law institution governing rented accommodation.

The institutions of land law may have proved ill designed for the needs of the rented accommodation market, although one should not exaggerate the problems they have generated. One remedy that the report does not consider, for the misfit between the institutions of the common-law and the requirements of housing law, is creation of a new institution, a new legal interest in land. This could be designed to meet the needs of the rented housing market. It would be a matter of technique whether this is done by deeming occupation agreements to be leases, or by recognising occupation agreements as legal estates in their own right. The problem of parallel institutions could be avoided, and the well-known structure of land law could operate when issues involving third party interests arose. Finally, it would be desirable if the new forms of occupation agreement recommended by the report, and indeed the new form of ownership recently created by statute (commonhold) were properly integrated into the corpus of land law. Land law may benefit from a new emphasis on the use value, as opposed to the transfer value, of land. The assumption that the institutions of land law must be taken as a given threatens to impair not only the development of housing law, but of land law as well.

GRAHAM FERRIS*

[2] See G Ferris and C Williams, "A Fruitful Parent of Injustice: Unilateral Service of Notice to Quit by a Joint Tenant" (2004) 3 Web JCLI.

*Senior Lecturer in Law, Nottingham Law School.

# RISK, CONSENT AND RESPONSIBILITY FOR OUTCOME

*Chester v Afshar* [2004] UKHL 41 (HL)

*(Lord Bingham, Lord Steyn, Lord Hoffmann, Lord Hope, Lord Walker)*

## INTRODUCTION

In medical law generally, and particularly for disclosure cases,[1] causation has often proved to be the major barrier to a successful claim.[2] Disclosure cases arise when the doctor fails to mention a risk that subsequently materialises and it is this omission that grounds the doctor's fault, rather than in any operative error directly resulting in damage. In order to establish causation, the claimant must show that she would have refused consent and not undergone the procedure. Where it is clear that had the claimant been warned she would not have gone ahead with the operation, the claim is straightforward. Similarly, where the claimant had a life-threatening condition, such that there was no real option but to proceed with surgery, the claim is bound to fail.[3] The answer is less clear where the claimant may have undergone the operation, but on a different occasion or with a different surgeon. In *Chappel v Hart*,[4] the High Court of Australia decided, by a majority, that causation would be satisfied by a postponement. The House of Lords has now, for the first time, considered the issue.[5]

## THE CASE

In *Chester v Afshar*, the claimant, who suffered severe back pain, was advised that surgery was her best option and she was referred to Mr Afshar, "an eminent neurosurgeon". Although she was reluctant to undergo surgery, Mr Afshar reassured her sufficiently for the claimant to give consent. Unfortunately, following the surgery she was left with severe neurological deficit in, particularly, her left leg. At trial, she alleged that Mr Afshar had failed to disclose the risk of nerve damage or paralysis and that, had he done so, she would at least have sought second and third opinions before deciding whether to undergo the operation. Mr Afshar claimed that he had warned her of the risks but the trial judge preferred the claimant's version and awarded damages for the failure to disclose. Mr Afshar appealed against the decision. The Court of Appeal held that the decision to at least postpone surgery was sufficient to establish causation and the doctor was liable in negligence.[6] The defendant appealed to the House of Lords, which, by a three to two majority,[7] rejected the appeal and upheld the Court of Appeal's decision.[8]

[1] See, *eg*, *Smith v Barking, Havering and Brentwood HA* [1994] 5 Med LR 285, 289 (decided 1988).

[2] See, *eg*, *Wilsher v Essex AHA* [1988] AC 1074, HL.

[3] *Chester v Afshar* [2002] EWCA Civ 724; [2003] QB 356, 376 at 38, CA.

[4] [1998] HCA 55; [1999] *Lloyds Law Reports Medical* 223; (1998) 72 ALJR 1344, HCA.

[5] *Chester v Afshar* [2004] UKHL 41.

[6] *Op cit*, n 3 above, 379 at [47].

[7] Lords Steyn, Hope and Walker found for the claimant and rejected the appeal. Lords Bingham and Hoffmann dissented.

[8] *Op cit*, n 5 above.

## THE "BUT FOR" TEST

The "but for" test is the most primitive test of causation. Notwithstanding McHugh J's judgment in *Chappel*,[9] however, it remains the starting point for determining causal responsibility.[10] Although it is widely acknowledged that this test is neither sufficient nor necessary to justify the attribution of responsibility,[11] the first step is to consider whether the claimant's injury would not have happened "but for" the defendant's breach of duty. In *Chester* the defendant's breach was the failure to disclose the risk of neurological damage. For this to have causally contributed to the claimant's injury, the failure to disclose must have affected the claimant's decision in some way. Traditionally, causation has been established on this test simply by the claimant showing that she would not have undergone the procedure "but for" the defendant's negligence. However, in *Chester*, the claimant was only able to say that she would not have had the operation there and then. She argued that, while she would have sought other opinion she was unable to say that she would not have undergone the operation.

Does this inability categorically to reject the operation affect the claimant's position? The undisclosed risk was quantified as being around 1–2%.[12] Most of their Lordships appeared to characterise this as a truly random risk with the consequence that if harm materialises it is pure chance. Lord Hoffmann, in his dismissively short dissenting judgment, which perhaps betrays as much by its brevity as by the proffered reasons, is most explicit in this by equating the situation to playing roulette.[13] Because the risk was random, it would remain the same whether or not the operation was delayed or performed by another surgeon. Thus, the patient would always face the same 1–2% risk irrespective of when the operation was performed. Because of this, Lord Hoffmann asserted that: "It follows that the claimant failed to prove that the defendant's breach of duty caused her loss. On ordinary principles of tort law, the defendant is not liable".[14] He drew this conclusion without any discussion of the normative aspects of causation, which is unfortunate since, as I will argue below, if the risk is truly random then postponing the procedure would satisfy the "but for" test. Lord Hoffmann's conclusion is only true if liability is determined on the basis of wrongful exposure to a risk. Focusing on risk rather than actual harm assumes a particular normative position (see below) regarding bad luck and responsibility for outcome and, if the "but for" test is satisfied then denial of liability should be justified.

In the present case, if the risk is truly random, the surgeon could fail to disclose a risk and in 98–99 times out of 100 the risk will not materialise; he will not be liable in negligence and he will have got away with his wrongdoing. This is a natural consequence of a small random risk. It would be both unjust and illogical to allow that same bias to work in his favour where the harm materialises but the patient is only able to claim that she would have postponed rather than avoided the procedure. Lord Steyn, in rejecting the defendant's appeal, correctly noted that:

> it is a distinctive feature of the present case that but for the surgeon's negligent failure to warn the claimant of the small risk of serious injury the actual injury would not have

[9] *Op cit*, n 4 above, at [25].

[10] *Op cit*, n 4 above, at [93] *per* Kirby J; n 3 above, 377 at [43].

[11] See, *eg*, Lord Hoffmann's much quoted judgment in: *South Australia Asset Management Corporation v York Montague Ltd* [1997] AC 191, [1996] 3 WLR 87, 94–95. See also *Fairchild v Glenhaven Funeral Services Ltd* [2002] UKHL 22; [2002] 3 WLR 89, HL.

[12] *Op cit*, n 5 above, at [5].

[13] *Ibid* at [31].

[14] *Ibid* at [32].

> occurred when it did and the chance of it occurring on a subsequent occasion was very small. It could therefore be said that the breach of the surgeon resulted in the very injury about which the claimant was entitled to be warned.[15]

Logically, if, on the balance of probability, the harm would not have occurred but for the wrongful failure to disclose, causation is effectively made out: just as an increase in risk is irrelevant to liability if no harm is caused, so it is irrelevant where it can be shown that that the defendant's wrongdoing caused the harm to materialise. It may be bad luck, but, assuming a random risk of 1–2%, the harm would have more likely than not failed to materialise if the operation had been performed on a different occasion, even if all other things remained the same. Thus, while causing the patient to consent to the operation at that time had no effect on the level of risk, it did cause her – on the balance of probability – to suffer the harm. Because the risk was less than 50% it is clear that, assuming a wholly random risk, the "but for" test would be made out and the defendant would be the single causal factor that was blameworthy and, therefore, justifiably liable for the damage.[16]

Although Lord Steyn noted that the breach of duty resulted in the injury, he then went on to imply that this was insufficient for liability on "established principles of causation".[17] This reflects a similar position adopted by all of their Lordships and what Stauch has described as the "generally accepted proposition" that, to satisfy legal causation, "the defendant's conduct must have created or added to the risk of the type of damage that subsequently occurred".[18] However, this is normative qualification of the "but for" test, justified on the basis of implicit assumptions concerning the notions of agency, autonomy, control and responsibility. It is better, for the sake of clarity, to make a clear distinction between establishing a factual causal association and determining responsibility and hence liability. It is submitted that Lords Steyn and Hope adopted the preferable approach in recognising the "but for" causal link and that Lords Bingham and Hoffmann were wrong to concatenate the normative arguments with the factual ones in holding that the "but for" test was not satisfied.[19]

Lords Bingham and Hoffmann were, as I suggested earlier, focusing too heavily on the risk rather than the damage that eventuated as a consequence: it is logically irrelevant that the risk would have been the same on another occasion if, on the balance of probabilities, the damage would not have occurred.[20] This focus on risk is perhaps a consequence of those cases where the "but for" test cannot be satisfied because of the presence of multiple tortfeasors or where the impact of the wrongdoing could not be sufficiently determined. In cases such as *Fairchild v Glenhaven*[21] and *McGhee v NCB*,[22] the courts appealed to a material change in the risk because the "but for" test left the blameless victim bearing both the consequences and the responsibility for the injury while allowing the wrongdoers to avoid liability. The same is not true of this case in which, while the claimant will be left bearing the consequences of the operation, the wrongdoer can be held responsible on the basis of the "but for" and remoteness tests.[23] The issue of whether the defendant has materially increased the risk is a red herring in respect of *prima facie* liability.

[15] *Ibid* at [19].

[16] See, *eg*, A Grubb, "Clinical Negligence: Informed Consent and Causation" (2002) 10 *Med L Rev* 322, 324

[17] *Op cit*, n 5 above, at [20].

[18] M Stauch, "Taking the Consequences for Failure to Warn of Medical Risks" [2000] 63 MLR 261, 264.

[19] *Op cit*, n 5 above, at 8, [29].

[20] *Op cit*, n 3 above, 377 at [40–41].

[21] [2002] 3 WLR 89, HL.

[22] [1972] 3 All ER 1008, HL.

[23] *The Wagon Mound (No 1)* [1961] AC 388, PC.

Of their Lordships, only Lord Walker approached the issue of risk in the medical context in a more sensitive way.[24] This may be because their Lordships generally felt constrained by the factual finding of the trial judge that postponing the operation would have had no effect on the level of risk faced.[25] As a term, risk is used in two very different contexts.[26] In one setting it is used to express a mathematical probability of, assuming a perfect system, the occurrence of a random event. The tossing of a coin or, to use Lord Hoffmann's example, the spinning of a roulette wheel, are examples. Alternatively, risk can be used to express the observed likelihood of an event materialising where some of the relevant factors, or their effects, are unknown or unquantified. This latter use of the term reflects its employment in the medical context where epidemiological statistics are relied on to make predictions about future events. This practice takes some observable, and hence controllable, characteristics, such as the patient's sex or age, in order to provide a degree of personal sensitivity to the quantification. However, since most individual characteristics, whether concerning the patient or professional, are not taken into account the figure is only a very rough guide.

The relevance of this type of statistic was appreciated by the House of Lords in *Hotson v East Berkshire HA*,[27] in which the court rejected the claim that, because 25% of similar injuries responded to treatment, the claimant in that case had been denied a chance of a successful recovery. In *Hotson*, the different outcome was attributable to the distinction between the two possible underlying mechanisms of injury. While superficially the injury appeared the same, in 25% of cases the disruption of the blood vessels was remediable, which would lead to complete recovery. In the other 75% of cases, the damage to the blood vessels was irremediable and disability was inevitable. Thus, because, on the balance of probabilities, the claimant could not prove he would have been in the 25% who responded to treatment he had not "lost the chance" of a successful outcome. The use of the risk statistic in *Chester* may be analysed in a similar way. The claimant might either be in the 98–99% of patients for whom such a risk would not materialise, or she could be in the 1–2% of unfortunates. There is strong evidence, from the fact that the risk did materialise, that the claimant was one of the unfortunate 1–2% of patients and thus the only way she could have avoided the harm would have been permanently to avoid the operation.

This approach would, contrary to the argument that Lord Walker seems to be advancing, support Lord Hoffmann's argument far more readily than the "random risk" approach. However, certain unknowns, such as the surgeon's characteristics or perhaps other operative personnel or even the environment, complicate the position. Common sense perhaps suggests that the patient's characteristics are the most relevant variables with the surgeon having the next greatest influence. As I have already argued, on the basis of the patient variables, the claimant may not have been able to show that she was not fated to suffer this injury. Appealing to the surgeon's characteristics also seems of little help as the defendant was acknowledged as one of the leading experts and the risk had not materialised in any of his previous patients.[28] Thus, recognising the non-random nature of risk may make the "but for" test harder to satisfy.

[24] *Op cit*, n 5 above, at [97].

[25] See, *eg*, n 5 above, at [7] *per* Lord Bingham.

[26] J Steele, *Risks and Legal Theory* (Hart Publishing, 2004) at 21–22. Steele relies on: I Hacking, *The Emergence of Probability* (Cambridge University Press, 1975).

[27] [1987] AC 750, HL.

[28] *Op cit*, n 5 above, at [46] *per* Lord Hope; at [95] *per* Lord Walker.

## THE NORMATIVE ARGUMENTS

While it is arguable that their Lordships' approach to the interrelationship between risk, harm and causation is flawed it is also, to a certain extent, a moot point. As I noted earlier, and as their Lordships were at pains to point out, the "but for" test both over-determines and under-determines the issue of responsibility.[29] One of the main reasons for constructing rules is to provide a guide that will encourage consistency and hence predictability. However, the goal of the court is to determine responsibility and hence legal liability. Rules help in establishing responsibility but they should only be used as a guide and not as an absolute determinant. As O'Neill argues:

> [r]ule-following . . . provides no criterion of "right" continuation: all rules are incomplete, and to "follow" them is to interpret them in a certain way. No rule can have written into it a determination of what it would be to follow it. Rules do not lay down complete answers.[30]

It is, therefore, the evaluative part of the judgments that are most relevant to the decision. This normative element is acknowledged by Lord Hope's appeal to policy,[31] and is further highlighted by Lord Steyn's comment on the relevance of the "but for" test. Lord Hope suggested: "[i]t can be said that Miss Chester would not have suffered her injury 'but for' Mr Ashfar's failure to warn here of the risks . . . But it is difficult to say that his failure was the *effective* cause of the injury".[32] It is the different normative evaluations that really distinguish the majority from the minority judgments.

The House of Lords could have argued for any of three different outcomes once the "but for" test had been dealt with: full recovery, recovery for the harm caused to the patient's autonomy or no recovery. I will discuss these options in turn.

### *Full Recovery*

This reflects the majority's decision and was justified by appealing to the whole basis for the doctor's duty to disclose. As Lord Steyn held: "[the claimant's] autonomy and dignity can and ought to be vindicated by a narrow and modest departure from traditional causation principles".[33] Similarly, Lord Hope suggested that, "the duty to warn . . . has at its heart the right of the patient to make an informed choice" and, if the appeal were allowed, it would strip the duty "of all practical force" and render it "devoid of all content".[34] This particular approach, however, while it seems straightforward enough and a reasonable justification, relies on an unspoken assumption about the nature of autonomy.

While there are many different views of autonomy,[35] the two approaches that appear most relevant to this case are:

(a) the more atomistic liberal view of autonomy as an essential element in the self-authorship of the individual's identity as moral agent; and

(b) the more socially-embedded view of autonomy supported by some feminists, communitarians and other anti-liberal views.[36]

[29] See, *eg*, *ibid*, at [18] *per* Lord Bingham.

[30] O O'Neill, *Towards Justice and Virtues* (Cambridge University Press, 1996) at 79.

[31] *Op cit*, n 5 above, at [85].

[32] *Op cit* , n 5 above, at [61], emphasis added.

[33] *Op cit*, n 5 above, at [24].

[34] *Ibid* at 86–87. See also Lord Walker at [101].

[35] See, *eg*, G Dworkin, *The Theory and Practice of Autonomy* (Cambridge University Press, 1988) at 6.

[36] For a feminist view see *eg*, L Barclay, "Autonomy and the Social Self" in C Mackenzie and N Stoljar (eds), *Relational Autonomy: Feminist Perspectives on Autonomy, Agency and the Social Self* (Oxford University Press, 2000) at 52; J Nedelsky, "Reconceiving Autonomy: Sources Thoughts and Possibilities" (1989) 1 *Yale Journal of Law and Feminism* 7.

The relevance of distinguishing these two conceptions of autonomy, which perhaps form a continuum lying between the two extremes of wholly individualistic and wholly socially embedded, is that they may provide different answers to the question whether responsibility for outcome should follow consent. As a legal and moral device, consent operates by permitting an otherwise forbidden act and so is morally and legally transformative.[37] There is nothing inherent to the concept of consent that requires responsibility for outcome to follow that consent. However, the liberal view of autonomy, which sees the person as the author of his self-determined choices, does support such a link between consent and responsibility. This would be consistent with a consumerist, free-market politic that sees consent as akin to contract between two equal competitors.[38]

The liberal approach is perhaps most evident in Lord Steyn's reliance on Ronald Dworkin's argument that autonomy "makes self-creation possible".[39] It is further supported by his emphatic claim that "a patient's right to an appropriate warning ought normatively to be regarded as an important right which must be given effective protection wherever possible".[40] Similar evidence comes from Lord Hope's reliance on Professor Jones' argument that "a patient with no rights is a citizen who is stripped of his or her individuality and autonomy, as well as her clothes, as soon as she walks into the surgery or hospital".[41] On this view, since the patient is not the author of the choice, although he or she has to bear the consequences, he or she is not responsible for it. This approach is also perhaps consistent with the political rhetoric of choice that is propounded by the present government.[42] Furthermore, it conforms with a less deferential approach to the medical profession and so reflects the trend that Lord Woolf argued, extra-judicially, was evident in recent legal judgments.[43]

This approach to autonomy, consent and responsibility for outcome has a perhaps unintended consequence if it is followed to its logical conclusion. In this case, the judgment was based on the finding that the claimant would have sought at least a second opinion and so have postponed the operation. The defendant precluded this by failing to disclose the relevant risk and so undermined her autonomy. If recovery were denied, this would undermine the law's protection of that autonomy, which was the whole purpose of requiring the duty in the first place. But, if the law is concerned with protecting this liberal view of autonomy, that highlights the importance of authorship and associates outcome responsibility with consent, then it should also allow recovery even where disclosure would not have altered the patient's decision.

This view of autonomy supports at least two reasons for disclosing a risk. First, the risk may be rationally relevant to the decision, and this may be so even where it would remain reasonable for the patient not to alter her decision because of the risk. Second, knowledge of the risk is required because it allows the patient autonomously to accept responsibility for that risk and allows the patient to be prepared should the risk

For a general discussion of the relationship between autonomy and the identity of the self see: MJ Sandel, *Liberalism and the Limits of Justice*, 2nd ed, (Cambridge University Press, 1998) especially 15- 65. For a more individualistic notion of autonomy see *eg* TL Beauchamp, JF Childress, *Principles of Biomedical Ethics*, 5th ed, (Oxford University Press, 2001) at 59.

[37] L Alexander, "The Moral Magic of Consent (II)" (1996) 2 *Legal Theory* 165.

[38] O O'Neill, *Autonomy and Trust in Bioethics* (Cambridge University Press, 2002) at 47.

[39] *Op cit, n* 5 above, at [18].

[40] *Ibid* at [17].

[41] *Ibid* at [58]; MA Jones, "Informed Consent and Other Fairy Stories" (1999) 7 *Med L Rev* 103, at 129.

[42] See, for example, the Department of Health's web site: http://www.dh.gov.uk/PolicyAndGuidance/PatientChoice/fs/en (accessed 27th October 2004).

[43] Lord Woolf's inaugural lecture in the Provost's Lecture Series, University College, University of London: "Are The Courts Excessively Deferential To The Medical Profession?" (2001) 9 *Med L Rev* 1.

materialise. For these reasons, the law should not require that disclosure would have altered the claimant's decision. This effectively makes the defendant an insurer against the bad luck that sees the risk materialise, an approach Lord Hoffmann rejected in his dissenting judgment.[44] The only justification for not following this through to its logical conclusion would be some countervailing policy argument.

*Recovery For Harm To Autonomy*

If recovery is to be restricted to recovery for the harm caused to the claimant's autonomy then the conception of autonomy adopted by the court must allow responsibility for outcome to be separated from consent. One possibility would be to adopt the second conception of autonomy as constrained by the social network in which each individual exists. In this view, the relationships within this social setting both make autonomy possible and constrain it. Without the support of the various co-existing communities we inhabit, choice would be severely restricted.[45] However, that support comes at the cost of the obligations we owe to the community. Furthermore, the influence of those relationships undermines the idea that we construct our own identity through our autonomous choices. Although autonomy may be valuable and worth protecting, it does not necessarily require an absolute connection between personal autonomy and responsibility. Such an approach may support an argument that the issue of outcome responsibility is more complex than simply tacking it on to the hem of consent's coat.[46] This would accord with the provision in the Declaration of Helsinki that holds the researcher responsible for any consequences, irrespective of the consent of the subject.[47] It may also underpin Lord Hoffmann's argument that while the doctor should not be responsible for the full cost of the injury "a modest *solatium*" might be paid because the "failure to warn ... was an affront to her personality".[48]

A parallel may be drawn with the majority's decision in *Rees v Darlington Memorial Hospital NHS Trust*.[49] In *Rees*, a wrongful pregnancy claim, the majority of the House of Lords held that the claimant could recover a conventional sum for the harm done to her reproductive autonomy, but not for the eminently foreseeable maintenance costs of raising a healthy child. This decision is justifiable as a response to the distributive justice quagmire that had developed as a consequence of the harshness of the House of Lords' decision in *McFarlane v Tayside HB*.[50] The decision in *Rees* to balance the responsibility for the outcome of the treatment event between the claimant and defendant bears similarity to Lord Hoffmann's argument for *solatium*. Given the added weight afforded to the right to a private life, which subsumes autonomy, by the passing of the Human Rights Act 1998, such an infringement would support awarding more than nominal damages.[51]

[44] *Op cit*, n 5 above, at [35].

[45] A McCall Smith, "Beyond Autonomy" (1997) 14 *Journal of Contemporary Health Law and Policy* 23, 37.

[46] See, for example, Roger Dworkin's argument that the action for informed consent should be eliminated: "Getting What we Should from Doctors: Rethinking Patient Autonomy and the Doctor-Patient Relationship" (2003) 13 *Health Matrix* 235.

[47] Declaration of Helsinki 1964 (as amended), paragraph 15. Available at: http://www.wma.net/e/policy/b3.htm (accessed 27 October 2004)

[48] *Op cit*, 5 above, at [33–34].

[49] [2003] UKHL 52, HL.

[50] [2000] 2 AC 59, HL.

[51] *Cornelius v De Taranto* [2001] EMLR, 12 at [66], HC. This aspect of the case was upheld on appeal without comment: [2001] EWCA Civ 1511; (2002) 68 BMLR 62.

*No Recovery*

If the socially embedded view of autonomy is biased slightly more towards the community, then, since consent and outcome responsibility have been divorced, it also supports the minority's position of not allowing recovery, unless the disclosure would have caused the claimant definitively to refuse to undergo the procedure. The position is complicated, however, because the claimant in *Chester* was a private patient.[52] The relationship between a doctor and his private patient fits more neatly with a consumerist ethic and more readily justifies the contractual approach to consent, risk disclosure and responsibility for outcome. However, the law generally does not distinguish between liability in private practice and liability in NHS practice.[53] Allowing the private patient to recover would create a general rule that would also allow the NHS patient to recover, which would plunder the NHS coffers, removing valuable funds needed for patient care. Although open to the criticism that NHS funds could be increased by raising or diverting tax revenue, the idea that one unfortunate patient should gain at the expense of the whole community of NHS patients, simply because her decision was not as well informed as it might have been, perhaps justifies the finding that the failure to warn was not a normatively relevant cause of her injury.[54] Rather, it was the unfortunate consequence of an operation performed primarily for her benefit. Since the operation was for her benefit, the loss should lie where it falls, as the impact of the consequences will be reduced through the social support provided by the NHS and other community or state-funded agencies.

Further support for the minority's approach comes from the view that healthcare generally is good for the community. It is, therefore, something that the community should support and protect. Recent events, such as the Bristol heart scandal, the organ retention scandal and the Harold Shipman affair, have threatened the confidence of the healthcare community. Perhaps this is reflected in Lord Bingham's conclusion that:

> The patient's right to be appropriately warned is an important right, which few doctors in the current legal and social climate would consciously or deliberately violate. I do not for my part think that the law should seek to reinforce that right by providing for the payment of potentially very large damages by a defendant whose violation of that right is not shown to have worsened the physical condition of the claimant.[55]

This is compounded by the argument that the NHS is facing a litigation crisis, with a huge bill threatening the provision of patient care.[56] The spectre of litigation also threatens to erode the rewards that might otherwise be gained from being a healthcare professional. This in turn will reduce the number of people wanting to enter the profession and increase the current staffing problems. This would be to the detriment of the community as a whole, including the unfortunate claimant. Since other means of support exist to ease her burden, it is better to see the injury as one of bad luck rather than any one individual's personal responsibility.

[52] *Op cit*, n 5 above at [37].

[53] See *Thake v Maurice* [1986] QB 644, 679 *per* Kerr LJ, CA.

[54] This is the same argument that Professor Jones thought justified the decision to reject the claim for compensation in the wrongful pregnancy case, *McFarlane v Tayside Health Board* [2000] 2 AC 59: MA Jones, "Bringing up Baby" (2001) 9 *Tort Law Review* 14, 19.

[55] *Op cit, n* 5 above, at [9].

[56] V Harpwood, "The Manipulation of Medical Practice" in M Freeman, A Lewis, (eds), *Law and Medicine: Current Legal Issues Volume 3* (Oxford University Press, 2000) at 47–48; L Mulcahy, "Threatening Behaviour? The Challenge Posed by Medical Negligence Claims" in M Freeman, A Lewis, (eds), *Law and Medicine: Current Legal Issues Volume 3* (Oxford University Press, 2000) 81, 83. Although Mulcahy commented that her research suggests the threat was exaggerated, in 2001, the National Audit Office (NAO) reported a seven-fold increase in costs since 1995–1996: NAO, *Handling Clinical Negligence Claims in England* (London 2001), 1 at: http://www.nao.gov.uk/publications/nao_reports/00-01/0001403.pdf (last visited 27th October 2004).

## CONCLUSION

In *Chester v Afshar,* the House of Lords has considered, for the first time, whether the decision to postpone a procedure has sufficient normative power to justify attributing causal responsibility in failure to warn cases. Recognising the limitations of the "but for" test, the House of Lords, by a three to two majority, held that it did. In this discussion of the case, I have focused on two aspects of the case. First, I suggested that the approach to risks was flawed and did not justify Lords Bingham and Hoffmann's conclusion that the failure to disclose had not caused the harm. However, I then argued that, because the "but for" test is both over- and under-determinative, this was of less importance than the distinct normative approaches taken in the majority and minority judgments. I suggested that the majority decision might be justified by adopting a liberal conception of autonomy that requires responsibility for outcome to follow consent. The minority judgments, which were less fully explicated than the majority judgments, are better supported by the anti-liberal conception of a socially embedded autonomy that divorces consent from responsibility for outcome. I also discussed a third approach (flirted with by Lord Hoffmann), which adopts a middle ground by allowing recovery for the infringement to patient autonomy.

Following *Chester,* the law is left with a relatively clear rule that it will be sufficient to establish causation in failure to warn cases provided the patient would have at least postponed the operation. If the autonomy argument is accepted, the logical consequence of this is that the patient should also be able to recover damages if the undisclosed risk materialises even where disclosure would not have altered her decision. This conclusion is perhaps supported by Lord Steyn's statement, which I quoted earlier, that the patient's right to be warned "must be given effective protection whenever possible".[57] This is, unfortunately, inconsistent with the more socially sensitive approach taken by the House of Lords in the wrongful pregnancy cases, culminating in *Rees v Darlington.* It does, however, accord with the dominant government rhetoric of choice and with the prevalence given to individual rights in the recent High Court case, *R (on the application of Burke) v GMC.*[58]

ALASDAIR MACLEAN*

[57] *Op cit,* n 5 above at [17].

[58] [2004] EWHC 1879.

*BSc (Hons), MBBS, PG Dip Law, M. Jurisprudence. Lecturer in Medical Law, The School of Law, the University of Glasgow.

# CAUSATION AND CONFUSION IN RESPECT OF MEDICAL NON-DISCLOSURE

*Chester v Afshar* [2004] UKHL 41

*(Lord Bingham of Cornhill, Lord Steyn, Lord Hoffmann, Lord Hope of Craighead, Lord Walker of Gestingthorpe).*

In *Chester v Afshar* the House of Lords was faced with an interesting problem of causation in the context of a "non-disclosure" case in medical negligence, *ie* a case where the claim is not that treatment itself was faulty, but that the doctor failed, at the earlier stage of obtaining consent, to inform the patient of a significant (and inevitable) risk in relation to it. In effect the problem was this: if the patient has the treatment and is injured when the risk in question materialises, what should the legal result be where the evidence shows that, had the patient been informed of the risk, he would have delayed the treatment (in order to seek a second opinion) but would eventually have had it anyway? In such a case, assuming that the risk of injury arising from the hypothetical delayed treatment would have been exactly the same as from the actual treatment, the doctor's failure to disclosure the risk has merely altered the time of the patient's exposure to it: it has neither created nor added to such risk. Should the doctor nonetheless be held liable to the patient for his injuries?

## THE FACTS

The claimant, Miss Chester, had suffered from progressively worsening back pain for a number of years when she was referred to the defendant, a distinguished consultant neurosurgeon, in November 1994. Although she was apprehensive about operations in general, Mr Afshar swiftly won her round to spinal surgery as offering the best solution to her condition and carried this out a few days later. Unfortunately, despite the exercise by the defendant of all due care in performing the surgery, a known risk of "cauda equina syndrome" (assessed at around 1–2 *per cent*) materialised, leaving Miss Chester with serious disabilities.

At trial, Mr Afshar acknowledged that it had been his duty, in line with good medical practice, to inform the claimant of this risk, and indeed he claimed to have done so. However, the judge preferred Miss Chester's account that nothing had been said to her about it and found Mr Afshar in breach of his duty of care under the *Bolam* test.[1] The judge then went on to consider the question, relevant to causation, of "what would have happened if the defendant had performed his duty?" On this, he accepted Miss Chester's evidence that, had she known of the risk, she would at the very least have postponed the surgery (so as to seek further advice as to whether it was really necessary), though she could not say that she would never have submitted to it at any future point. This, in the judge's view sufficed to establish causation and liability: after all, if the claimant had had the surgery on some later occasion, the risk – which, as noted, was in the range of 1–2 *per cent* – would very probably not have materialised.

The Court of Appeal (Hale LJ, Sir Christopher Slade and Sir Denis Henry)[2] dismissed the defendant's appeal on both breach and causation, and by the time the case reached the House of Lords only causation remained in dispute. It was that issue, as sketched at the outset of this note, that gave rise to greatest difficulty and – as we

[1] See further the discussion in the text below, at n 4.

[2] [2002] EWCA Civ 724; [2003] QB 356; the Court of Appeal delivered a single judgment in the case.

shall see – provoked a three:two split among the Law Lords. However, also of great interest from the medico-legal point of view is what the House of Lords had to say on the anterior question of breach of duty in relation to non-disclosure of medical risks. Even though its remarks in this regard are *obiter*, this was the first time since the *Sidaway* case[3] nearly 20 years ago – a decision long subject to academic criticism for its apparent failure to uphold patient rights – that the House of Lords has had occasion to touch upon this important matter.

## THE BREACH OF DUTY – WHERE DOES *CHESTER* LEAVE *SIDAWAY*?

In *Sidaway*, the House of Lords extended the application of the *Bolam* test[4] – used in cases of alleged negligent diagnosis and treatment – to the issue of breach in relation to a doctor's failure to disclose risks while obtaining consent. This test privileges accepted medical practice by holding that a doctor whose actions are endorsed by a responsible body of his or her peers will not (or, at any rate, only very rarely[5]) be in breach of duty: applied to the disclosure of risks, it meant that a doctor typically had to disclose only those risks that other responsible doctors would deem proper to disclose. Subsequently, it became common for academics to contrast this paternalistic, "prudent doctor" stance with the more pro-autonomy, "informed consent" approach adopted in various US states, as well as Australia and Canada,[6] which requires the doctor to disclose all risks that a reasonable patient would wish to know about in deciding whether to consent to treatment.

In *Chester*, the House of Lords – without expressly doubting *Sidaway* – acknowledged that in the intervening years the law in England too has moved towards recognising the patient's right to know of significant risks to the treatment the doctor is proposing.[7] In his speech, Lord Steyn specifically embraced the language of "informed consent", noting that, "[t]he court is the final arbiter of what constitutes informed consent. Usually, informed consent will presuppose a general warning by the surgeon of a significant risk of the surgery".[8] His Lordship approved Lord Woolf MR's *dictum* in *Pearce v United Bristol Healthcare NHS Trust*, that ". . . if there is a significant risk which would affect the judgment of a reasonable patient, then in the normal course it is the responsibility of a doctor to inform the patient of that significant risk".[9]

For his part, Lord Hope, who was the only other member of the House to address the issue in detail, was more cautious, and preferred to see in the *Sidaway* decision itself an approach aleady sufficiently protective of patient autonomy. As he noted:

> Lord Templeman said [in *Sidaway*] that he did not subscribe to the theory that the patient is entitled to know everything. Some information might confuse and other information might alarm the patient. So it was for the doctor to decide in the light of his training and experience what needed to be said, and how it should be said. But he went on to add these

[3] *Sidaway v Board of Governors of the Bethlem Royal Hospital and the Maudsley Hospital* [1985] AC 871.

[4] From the case of *Bolam v Friern HMC* [1957] 1 WLR 582.

[5] See *Bolitho v City & Hackney HA* [1998] AC 232.

[6] See *Canterbury v Spence* (1972) 464 F 2d 772; *Rogers v Whitaker* [1993] 4 Med LR 79; *Reibl v Hughes* [1980] 2 SCR 880.

[7] See Lord Walker [2004] UKHL 41, at para 92: "during the twenty years which have elapsed since *Sidaway* the importance of personal autonomy has been more and more widely recognised".

[8] Para 14.

[9] [1999] PIQR P53, at P59, cited at [2004] UKHL 41, para 15. However, Lord Steyn acknowledged the possible defence of "therapeutic privilege" to justify non-disclosure where the doctor reasonably believes the information will have a serious negative impact upon the patient's mental well being. As he noted, this was not relevant on the facts.

> words ... "At the same time the doctor is not entitled to make the final decision with regard to treatment which may have disadvantages or dangers. Where the patient's health and future are at stake, the patient must make the final decision."[10]

Lord Hope went on to suggest that, "the right to make the final decision and the duty of the doctor to inform the patient if the treatment may have special disadvantages or dangers go hand in hand".[11]

Significantly, as the House of Lords recognised, nowadays good practice guidelines promulgated by doctors' own professional organisations make clear the rights of patients to be informed of significant treatment risks.[12] In the light of this, it could be argued that the dichotomy previously identified, between the *Sidaway* "prudent doctor" and US-style "informed consent" approaches, has disappeared. This is because, on the logic of the *Bolam* test itself, if responsible accepted practice generally is to disclose risks that a reasonable patient would wish to know about, the failure of a particular doctor to do so in a given case will be a straightforward breach of duty.

## CAUSAL CONNECTION – AN UNFORTUNATE ELISION

As mentioned earlier, the principal issue at stake in *Chester* – and the matter on which the defendant's appeal turned – was causation. At this juncture, we should remind ourselves that, as regards the legal inquiry into liability for damage, causation falls into two distinct parts, namely *factual* causation and *legal* causation.[13] The former is a matter of causal necessity (normally tested by the "but for" test): "was the defendant's wrongful conduct a necessary condition for the claimant's injury?" By contrast, the legal causation question concerns whether, notwithstanding its status as a necessary condition for the harm, the defendant should be excused liability for his faulty conduct on the basis that – while reproachable for some other reason – such conduct did not foreseeably create or add to the risk of the injury that actually occurred. By failing to keep this distinction in mind, and referring simply to "causation" or "causal connection" as shorthand for both, the House of Lords in *Chester* made life unnecessarily difficult for itself.

It ought to be clear that, in *Chester*, *factual* causation was present: Mr Afshar's breach, in failing to advise of the risk, had changed the course of events and led to the claimant being injured when, otherwise, she would almost certainly have escaped such injury. As Lord Steyn commented, "... but for the surgeon's negligent failure to warn the claimant of the small risk of serious injury the actual injury would not have occurred when it did and the chance of it occurring on a subsequent occasion was very small".[14] However, surprising though it may seem, precisely this point was denied by the minority (Lords Bingham and Hoffmann). Thus Lord Bingham suggested that the fact that the injury was as liable to occur whenever Miss Chester had the surgery (a point going to legal causation) meant that (factual) causation, in terms of the "but for" test, was not satisfied.[15] In similar vein, Lord Hoffmann began his speech with the following argument:

[10] Para 54.

[11] Para 55.

[12] Lord Steyn referred, at para 26, to the Royal College of Surgeons' guidance, *Good Surgical Practice*. See also Lord Hope's remarks at para 58.

[13] The latter is often also referred to as "remoteness of damage".

[14] Para 19.

[15] Para 8. His Lordship went on to suggest that if recovery were allowed in *Chester*, it would logically also have to be allowed in a case like *Smith v Barking, Havering and Brentwood HA* [1994] 5 Med LR 285. However, the facts of the latter

> The purpose of a duty to warn someone against the risk involved in what he proposes to do, or allow to be done to him, is to give him the opportunity to avoid or reduce that risk. If he would have been unable or unwilling to take that opportunity and the risk eventuates, the failure to warn has not caused the injury. It would have happened anyway.[16]

With great respect, this is not sustainable. If the "it" in the last sentence refers to the injury, then the proposition is false. As already noted, it was very clear on the facts of *Chester*, that, if the warning had been given, the injury to the claimant would – in all likelihood – not have occurred. What is true is that she would in any event have been exposed to the same *risk* of injury. But the risk of a thing is not the same as the thing itself: by definition a risk may often not materialise. On the facts of this case, the risk of injury to Miss Chester – had surgery been delayed – was 98–99 *per cent* likely not to materialise.

Admittedly, *prior* to the actual operation in which the risk did unfortunately materialise, one would also have said that it was 98–99 *per cent* likely not to do so. Analytically, the reason for this is that, apart from the surgery itself, further independent causal factors had to be present for the injury to occur (were it not so, the surgery – however well-performed – would *invariably* result in injury), and those factors appeared, as the statistics showed, only in 1–2 *per cent* of cases. However, while it follows that the surgery executed by the defendant was not sufficient for the injury to Miss Chester, it was indisputably *necessary*: it was one element in a causal set (also containing those additional, more or less coincidental factors) that produced the injury.[17] And if Mr Afshar had properly discharged his duty (by warning the claimant of the risk), the fateful surgery would not have taken place.

## THE SIGNIFICANCE OF FACTUAL CAUSATION

The reason why, once established, factual causation is significant is two-fold. First it prevents the claimant from getting a windfall. It seems right (as the minority in Chester suggested, while failing to see that the claimant fulfilled the relevant condition[18]) to reserve substantial damages for cases where the doctor's conduct has altered the course of events and resulted in injury (when otherwise there would very likely have been no injury). Indeed, this is merely an instance of the general requirement, foundational in the tort of negligence, that the claimant must show "damage".

As an aside, it could be argued that by addressing medical non-disclosure cases in negligence, the law gives imperfect expression to autonomy: the sanction for the breach of the patient's right to information applies only in those cases where he can show that he would positively have acted upon the information in issue.[19] However, it is suggested that this problem is not acute. In the first place, some damages – for mental distress – may in fact be awarded even where the patient's prior knowledge of the risk would have made no difference to his decision to submit to immediate treatment. He or she will after all have suffered the shock of suffering a sudden, and for him or her

case (a High Court decision from 1989) differ from those of *Chester* precisely in that the evidence showed the claimant would have submitted to the same surgery *then and there* even if she had known of the risk: *ie* factual causation was *not* made out.

[16] Para 28.

[17] For a discussion of causal set analysis and how it can illuminate complex problems of factual causation and risk, see Stauch, M, "Causation, Risk and Loss of Chance in Medical Negligence" (1997) 17 *OJLS* 205.

[18] See in particular Lord Bingham's remarks at para 9.

[19] If, by contrast, the law positioned such cases in the tort of battery (by finding that the non-disclosure of significant risks vitiated consent), the patient's hypothetical response to disclosure would be irrelevant.

unexpected, injury as a result of the treatment.[20] Secondly, the law here is supplemented by other regulatory mechanisms that also serve to encourage respect for patient autonomy, notably the disciplinary codes promulgated by professional bodies such as the GMC.

Be that all as it may, the point that factual causation must be present before the law will fully vindicate a patient's right to be told of risks is of course no argument for denying redress where such causation has been established. Indeed, the second, more positive, reason why proof of factual causation is significant is that it gives rise to a powerful prima facie case for compensation: "if you had behaved properly, I would not have been injured!" Qualifications on liability based upon remoteness/legal causation points are relatively unusual. This is all the more so in cases where the injury intrudes in close temporal proximity to the wrongful act, and was readily foreseeable (both of which were true here).

The legal causation issue that nonetheless arose in *Chester*, of whether the law should acknowledge – by a no-liability finding – that Mr Afshar's breach had exposed Miss Chester to a risk she would have been exposed to in any event, will be looked at shortly. However, before doing so, it is worth briefly addressing two problems of quantum (closely linked to the issue of factual causation), which may occur in this type of case.

## QUANTUM ISSUES

Suppose, first of all that, though identifiable in advance only as a small risk, it subsequently becomes clear that the factors that, together with (carefully executed) surgery, led to the patient's injury were inherent in the latter's constitution: *ie* he was predisposed to suffer that injury.[21] In such a case, the doctor's wrongful failure to advise of the risk, which brought forward the surgery, has simply accelerated an injury that would have occurred later in any event. It is true that factual "but for" causation is still present: the patient would not have suffered the injury *when he did*, if the doctor had given the warning. Nevertheless, his loss – apart from mental distress – only extends to a short period of freedom from the injury in question. This fact should be recognised, in accordance with normal quantum principles, by a reduced award of damages.[22]

In *Chester*, the above issue did not arise: the risk of caudal equine syndrome appears to have been "free-floating": as liable to strike one patient as the next. However, the second quantum problem does so, albeit in a relatively minor form.[23] The problem is this: to what extent should the fact that the claimant would ultimately have run the same risk of injury also be reflected in discounted damages? In *Chester* itself, any such discount would only be in the order of 1–2 *per cent*. However, another case could quite easily arise in which the risk attached to the hypothetical delayed surgery is very much

[20] An award of this form was made in *Smith v Barking, Havering and Brentwood HA* (*op cit*, n 15).

[21] It is easy to imagine that, in some cases, a risk quotient of *eg* 1–2 *per cent*, attaching to a given medical procedure, is referable to the fact that 1–2 out of every 100 patients have the predisposition in question, but it is not possible in advance to say who they are.

[22] The same would also be true if, without the (injurious) treatment, the patient was likely to have suffered significant injury as the result of the progress of his underlying condition. These points were recognised by the Court of Appeal in *Chester*: see [2002] EWCA Civ 724, at paras 38 and 41–42.

[23] Damages in *Chester* have yet to be assessed (the matter having been set aside by the trial judge pending resolution of the liability issue).

greater, say 30 *per cent*. Here, it seems right for damages to be reduced to take account of the substantial chance that the patient would later have suffered the same injury anyway.[24]

## THE MAJORITY APPROACH TO (LEGAL) CAUSATION

Once the issue in *Chester* is seen clearly for what it is, it is submitted that difficulties in finding in favour of the claimant largely fall away. As noted earlier, the presence of factual causation already produces a significant dynamic in favour of liability. The further question, as to how far the normal legal causation requirement that the defendant's breach created or added to the risk, should be relaxed in this type of case had been exhaustively discussed in *Chappel v Hart*,[25] a recent case from the High Court of Australia with similar facts, as well as in the meticulous judgment of the Court of Appeal in *Chester* itself.[26] In fact there would seem to be good reasons for such a relaxation. Most importantly, as the majority in the House of Lords recognised, the doctor's duty to inform patients of risks would otherwise be very much attenuated. As Lord Hope commented:

> To leave the patient who would have found the decision difficult without a remedy, as the normal approach to causation would indicate, would render the duty useless in the cases where it may be needed most. This would discriminate against those who cannot honestly say that they would have declined the operation once and for all if they had been warned. I would find that result unacceptable. The function of the law is to enable rights to be vindicated and to provide remedies when duties have been breached.[27]

At a more general level, we should recall the reason for normally insisting, at the legal causation stage, that the defendant's conduct increased the risk of the claimant's injury. This is so as to rule out liability for coincidences of the sort canvassed by Lord Walker in his speech, such as where a passenger in a speeding taxi is injured by a falling tree.[28] In such cases it is true that, had the defendant behaved properly, the injury would have been avoided. Nonetheless, the rationale for the rule wrongfully breached by the defendant (namely, to reduce the risk of injury over a range of similar cases) is confounded by the particular circumstances at hand: the injury would here have been just as likely to occur if the rule had been respected.

Importantly, the above point has only limited force in medical non-disclosure cases:[29] here the rule of conduct (requiring disclosure of significant treatment risks) is designed not principally to reduce the risk of injury from treatment: in many cases – including *Chester* – the risk in question cannot be reduced. Rather, the main function of the rule is to promote the distinct goal of the patient's autonomy. The patient should be told

[24] In *Chappel v Hart* (1998) 72 ALJR 1344, Kirby J, in the High Court of Australia, suggested (at para 93-10) that a discount should occur, unless the residual risk was so small as to be merely speculative. In England too it is standard practice for damages to be discounted in line with the claimant's chance of suffering the same injury later in time: see *eg Smith v Leech Brain* [1962] 2 QB 405.

[25] (1998) 72 ALJR 1344.

[26] *Op cit*, n 2.

[27] Para 87. See also Lord Steyn's comments at paras 24–25.

[28] Para 94. (It is of course being assumed that the probability of the tree falling is unaffected by the speed of the passing vehicle).

[29] It would arguably have force in a case where the patient can show both that the doctor breached his duty by failing to advise him of a given risk (Risk A), and that, had he known of it, he would have declined treatment, but where his injury resulted from the occurrence of a different risk (Risk B), and the doctor either was not in breach by failing to disclose the latter or had in fact disclosed it. The remarks of Lord Walker (*ibid*) support the exclusion of liability here.

that there is an inevitable risk to the proposed treatment precisely so that he can decide if he wishes to submit to it and, if he does so decide, make contingency plans in case it should materialise.

## CONCLUSION

It is suggested that the majority of the House of Lords was quite right to find liability in this case: indeed, any other conclusion would have been a major disappointment. The surprise is rather that the two very eminent law lords who dissented should have become confused on the causation issue. In this regard, *Chester* should operate as a powerful reminder of the need, analytically, to keep the issues of factual and legal causation separate. At the same time the proof issue in respect of the former has been clarified: the patient need only show that they would probably have delayed the proposed treatment, not that they would have refused it for all time. Nonetheless, this does not represent a new concession to patient rights, but was inherent in the logic of the "but for" test all along. Insofar as any increase in the liability of doctors could result – and that seems to have been an underlying concern of the minority[30] – it is submitted that this can be adequately controlled (in addition to the control already presented by factual causation) by considerations at the quantum stage.

MARC STAUCH*

[30] See Lord Bingham's comments at para 9.

* MA (Oxon.), Solicitor, Centre for British Studies, Humboldt-Universität zu Berlin.

NOTTINGHAM LAW JOURNAL

VOL 14(1) 2005

# BOOK REVIEWS

*Book reviews and books for reviewing should be sent to the address given at the beginning of this issue*

## WOMEN IN THE LEGAL PROFESSION

*Women in the Law: strategic career management* General Editor ELIZABETH CRUICKSHANK, London, The Law Society of England and Wales, 2003, xvi + 334 pp., Paperback, £29·95, ISBN 1-85328-822-5

It is implicit that this is a text intended for young women (mature entrants, at least those post childbearing, are not really considered) considering taking up a career in the law. It should be said at the outset that one of its contributors, Sally Woodward, is a professor of the Nottingham Law School (for completeness, and the avoidance of confusion, the Richard Payne responsible for section 1.3 is not our own Richard Payne).

Despite the apparently intended audience, several of its chapters are of more relevance and appear to be directed to those already in employment and some have no specific "legal" bias. Perhaps the idea is that the book is for law undergraduates or CPE/GDL students and intended to be kept by them during the early part of their careers. At £29.95, it might be rather expensive for the student market and, of course, given Clementi and more particularly the Law Society's Training Framework Review, the advice given may now have a fairly short shelf life at least insofar as the solicitors' profession is concerned. One remains, as will be apparent throughout this review, somewhat confused as to the actual market for the text, intrinsically interesting as it might be.

The book is divided into three sections: Part One, Background; Part Two, Interviews and Part Three, Advice.

Following a short history of women in the solicitors' profession and at the Bar, the remainder of Part One consists of a section outlining "Careers in the Law for Women" and immediately proceeds to a chapter on "Taking control of your work-life balance" which, whilst of interest, is adapted from a chapter in another text without any obvious turn towards the legal professions. "Women in the Law", it seems from the career advice section, means solicitors, barristers, the judiciary and academics and does, therefore, serve to exclude the many women employed in other legal capacities, particularly the membership of ILEX (as at least one of the interviewees in Part Two is a "five years' woman", this might be considered an odd omission). The careers advice section is brief and one cannot help feeling that less generalised information would be available elsewhere. For example, a rapid survey of the faculty of this Law School and that of the University of Nottingham reveals that the possession of a PhD, whilst no doubt useful, particularly in research-based activity, is hardly an absolute pre-requisite for an academic career in law[1] particularly if one wishes to establish an

academic career in the BVC or LPC. It should also be pointed out for clarity that, as a publication of the Law Society of England and Wales, Scotland and Northern Ireland are omitted from discussion.

Part Two was, for this reviewer, the most interesting section of the book, and, as a social history, is significant. The general editor reports interviews with 28 women in the law: including 16 solicitors and seven barristers as well as academics, Commonwealth lawyers and a novelist. Whilst the text itself acknowledges that the majority are from professional or middle-class backgrounds, by no means all are. Only one interviewee appears to be of a non-white ethnic minority. Each interviewee, including for example, Margaret Brazier, Harriet Harman, Barbara Mills and Eileen Pembridge, gives an outline of her career, her reasons for entering the legal profession and generally some advice for new entrants to the legal professions. Given the rapid changes both in the professions and in legal education, it might have been helpful to put each interview into its historical context. As far as I can establish, three interviewees qualified in the 1950s, three in the 1960s, 13 in the 1970s and three in the 1980s. The others I cannot tell and the currency of their experiences in and immediately prior to qualification can, in some cases, only be divined from guessing the age of the interviewee from her photograph. Whilst the thoughts and advice of senior women in the profession, taken from a wide range of types of practice, are of huge value, if the book is indeed targeted at the current undergraduate (or even the A-level student facing top-up fees), the experience of those of more recent qualification might also have been useful.

Part Three is, then, the section which the trainee or newly qualified might wish to keep by her. Advice is given on soft skills, managing time, dealing with stress, leadership and "rainmaking and networking". All is potentially highly useful, although the sections on soft skills and time management do not address any issues specific to women and there is again a mixed message about the likely readership (*eg* page 272 "strategic placement of your name, image and expertise"; page 299 "a newly appointed practice group head") except in the piece on rainmaking and networking which does take an approach recognising the newly qualified lawyer.

Those of us who have been "women in the law" for any length of time will have our own anecdotes and memories of past or current inequalities. This writer, schooled in the late 1980s that the last thing a woman interviewee for a training contract should admit to was an ability to type (and recalling a former colleague who had qualified in the 1960s and was apparently assumed to possess a gene for typing, interviewers commenting "Well, you won't need a secretary, will you?") was amused by the assertion (page 11) that "Having keyboard skills will increase your chances of [obtaining a summer placement]".

Much has changed since the University of Queensland had a pass mark of 50% for male undergraduates; the pass mark for women was 75%.[2] Much has changed but, as the emphasis throughout this text on the work-life balance demonstrates, much still needs to change. Young women lawyers (and some of their older sisters) might usefully read this book. Whether – and at what stage of their careers – they will buy it remains a very different question.

JANE CHING*

[1] The University of Nottingham law faculty, on a review of its website, scores roughly half and half between PhDs and those with other qualifications (including foreign qualifications, LLMs and professional qualifications). Our, own, substantially more vocationally focused law school has roughly 11 PhDs out of a faculty of 138.

[2] See page 182.

* Reader, Nottingham Law School.

## CONSUMER LAW

*Blackstone's Guide to Consumer Sales and Associated Guarantees* by ROBERT BRADGATE and CHRISTIAN TWIGG-FLESNER, Oxford, Oxford University Press, 2003, xxvi + 309 pp., Paperback, £34·95, ISBN 0-19-925594-6

This book deals in detail with the effect of the Sale and Supply of Goods to Consumers Regulations 2002 ("the Regulations"). The Regulations implement EC Directive 1999/44/EC ("the Directive") on certain aspects of the Sale of Consumer Goods and Associated Guarantees, which gives additional rights to consumers who acquire goods. The book sets out in full the Directive, the Regulations and the statutes that have been amended. So the Sale of Goods Act 1979, the Supply of Goods and Services Act 1982, the Supply of Goods (Implied Terms) Act 1973 and the Unfair Contract Terms Act 1977 are reproduced in full, with the amendments made to them italicised. It is useful to have all of the affected legislation in one volume but nevertheless the Directive and the legislation amount to about 20 *per cent* of the book. Whilst academics, practitioners and students are likely to have copies of the statutes and the Regulations, and perhaps of the Directive, elsewhere, the authors hope that the book will also be useful to retailers, consumers and advisers.

The approach taken by the book is first to consider the provisions of the Directive, second to consider the corresponding rules of domestic law and only then to consider the requirements of the Regulations. The Regulations leave a number of questions unanswered and the authors explore deficiencies and ambiguities in depth, with an underlying theme that the time may well have arrived when two distinct codes of sales law are needed, one for business buyers and the other for consumers.

The first chapter begins by tracing at length the origins of the Directive and the history of the Directive's implementation into domestic law. It then considers how effective the implementation has been from two perspectives. First, whether the new legislation promotes consumer protection in a clear and coherent way. Second, whether the legislation has implemented the Directive sufficiently well to comply with the requirements of EC law. A considerable number of words are devoted to ascertaining exactly what these requirements are. Of the authors' intended audience, academics are likely to find this of more interest than the other groups at which? the book is aimed.

The main effect of the 2002 Regulations is to give consumers who have acquired goods that do not conform to the contract a hierarchical set of remedies. At the first level of the hierarchy the consumer is entitled to either repair or replacement of the goods. The second level remedies are reduction of the purchase price or rescission of the contract. The second level remedies are not available unless the first level remedy for which the consumer opted is either impossible or disproportionate in relation to another of the remedies. These new remedies are additional to any other rights that the consumer might have, such as the right to reject the goods for breach of a statutory implied term. This might, or might not, sound simple enough but as the authors have previously pointed out the new legislation gives rise to several questions to which it does not provide answers.

The authors approach the new rules on conformity with the contract by including separate chapters on the following matters: the scope of the legislation and its definitions; the requirement that goods must be in conformity with the contract; the remedies available; exclusion of liability and the position of third parties. In general, these chapters are divided into five or so bullet points. They have an introduction and a conclusion and in between deal with the provisions of the Directive, the position

prior to the implementation of the Directive and the way in which the Directive has been implemented. This formulaic approach is useful in allowing the reader quickly to ascertain where to look for relevant material.

Chapter 2, which deals with definitions, is particularly interesting in its explanation of the way in which the Unfair Contract Terms Act 1977 definition of a consumer, which the Regulations adopt, has been amended in order to implement the Directive.

Chapter 3 examines the requirement that the goods should conform to the contract. The authors first deal with the introduction of the Sale of Goods Act 1979, sections 14 2D – F, which set out the ways in which public statements about the goods can now be relevant circumstances in deciding whether or not goods are of satisfactory quality. Sale of Goods Act 1979, section 48F defines the meaning of conformity with the contract by stating that

> For the purposes of this Part, goods do not conform to the contract of sale if there is, in relation to the goods, a breach of an express term of the contract or a term implied by section 13, 14 or 15 above.

The domestic requirements of sections 13 and 14 are examined in considerable detail, as are the requirements of the Directive. However, I suspect that for most readers the most useful part of the chapter will be material that deals with UK implementation of the Directive.

The remedies available under the Regulations are surprisingly complex and in Chapter 4 the authors explore them in considerable detail, noting both theoretical and practical difficulties. Two tables at the end of the chapter give a useful overview of the various rights now available to a consumer. The first of these deals with the position where a consumer has bought goods that do not conform to the contract. The second deals with the remedial regime that applies to contracts to supply and install goods.

Chapters 5 and 6 deal respectively with exclusion of liability and the position of third parties. Both chapters discuss the domestic law position and the requirements of the Directive at considerable length. Article 7 of the Directive states that

> [a]ny contractual terms or agreements concluded with the seller before the lack of conformity is brought to the seller's attention which directly or indirectly waive or restrict the rights resulting from the Directive shall, as provided for by national law, not be binding on the consumer.

After a detailed consideration of the Unfair Contract Terms Act 1977 and the Unfair Terms in Consumer Contracts Regulations 1999 the authors conclude that Article 7 has not been properly implemented.

The second major effect of the 2002 Regulations is to make guarantees of consumer goods legally binding as a contractual obligation. This is considered in Chapter 6. The authors explore the somewhat uncertain position that applied before the Regulations became effective. They also deal at length with whether or not a consumer guarantee is legally enforceable by a person other than the consumer who bought the goods. This is a matter of considerable uncertainty as although the relevant regulation states that a consumer guarantee takes effect as a contractual obligation owed by the guarantor, it does not state to whom this obligation is owed.

The final two chapters deal with the liability of manufacturers and with the position of retailers. Manufacturers can become liable if they give a guarantee with their goods, something that they are not obliged to do. Retailers seem to have most to lose from the new Regulations. They remain liable to consumers for faulty goods and can now be liable for public statements about the goods made by the manufacturer. In addition, they may now be required to repair the goods, something that in many cases it would

surely be more appropriate for the manufacturer to do. The authors explore whether Article 4 of the Directive helps the position of retailers. It states that where a retailer is liable under the Directive he should be able to pursue remedies against the person or persons liable in the contractual chain. This seems to offer adequate protection, but the final sentence of Article 4 states that

> The person or persons liable against whom the final seller may pursue remedies, together with the relevant actions and conditions of exercise, shall be determined by national law.

Depending upon whether a broad or a narrow view of Article 4 is taken, it is arguable that the limited remedies available to retailers, as compared with the remedies that they must offer consumers, and the different exclusion clause regimes, mean that Article 4 has not been properly implemented into English law.

At first I was somewhat sceptical of the authors' claim that the book was intended to be useful to all those interested in consumer law, including academics, practitioners, retailers, consumers and their advisers. The book will certainly be of interest to academics. It explores a wide range of academic issues in a detailed but readable style. Practitioners should also be able to make use of the book. However, it would seem likely that most practitioners would be interested primarily in what the legislation now states, rather than with academic debate about whether or not the Directive has been properly implemented. The bullet-pointed structure and the reproduction of the relevant legislation allow the reader to quickly find the relevant material. The book might prove a little too much for consumers, retailers and for many of their advisers. However, it is possible to envisage that the two tables set out in Chapter 4 might be a starting point, enabling them to discover if any rights exist and if so the conditions which must be satisfied. One audience, which the authors did not mention, was students. Few law courses would be likely to require the depth of knowledge of the new Regulations that this book provides. However, much of the book sets out the general position in English law as it relates to the statutory implied terms, the Contracts (Rights of Third Parties) Act, remedies and the ways in which remedies can be lost. The book is up to date and erudite in its treatment of these subjects. The more serious students of sales law might well find this book a very useful addition to their main textbook.

EWAN MACINTYRE*

* Senior Lecturer in Law, Nottingham Law School

NOTTINGHAM LAW JOURNAL

VOL 14(1) 2005

# NOTTINGHAM MATTERS

*This section documents major developments and research projects within Nottingham Law School together with responses to public consultation exercises and other public contributions made by its staff.*

## HEALTH LAW TEACHING IN THE LAW SCHOOL

The problems of health law are legion. Almost every day stories appear in the media about clinical negligence or some innovative new medical technique that pushes back the frontiers of medicine, whilst raising fundamental ethical and social issues. Health law is a modern and contemporary academic subject, which demands that its teachers and students keep up to date with issues. Its contemporary nature has gained it many fans, none the least in our law school.

Health Law is a major feature of the curriculum at Nottingham Law School and has been for many years. We currently have ten teaching staff who teach on health law programmes. A significant number of these staff also maintain health law research specialisms. Health law is also popular with our undergraduates. Medical law is offered as a third year LLB option subject and has been the most popular option for students for a number of years. In the 2004–2005 academic year out of 266 students on the final year of the LLB, 123 chose this option.

We also offer a LLM in Health Law, which is in its third year of operation. Recruitment is around the mid-30s with students coming from a variety of backgrounds: medicine, nursing and also directly from undergraduate studies. A number of international students have joined the programme, some from as far as South America. We offer the course on a full-time and part time basis. The subject options on the LLM are currently:

(a) Introduction to Law;
(b) Legal Research Writing;
(c) Introduction to Health Law;
(d) Medical Ethics;
(e) Medical Malpractice Litigation: In Context;
(f) Mental Health Law;
(g) Treating the Incapable Patient;
(h) The Employer and Health Law;
(i) The Law Relating to Organ Transplants;
(j) Ethical and Legal Issues in Human Reproduction;
(k) Medicine, Pharmacy and the Law;
(l) End of Life Issues.

Health research organization ESCI kindly sponsors an annual prize for the student with the best academic profile on the course: with this year's resultant ESCI achievement award going to distinction student Lara Kretzer.

The director of the LLM programmes is Kay Wheat, Reader in Law and her deputy is Austen Garwood-Gowers, Senior Lecturer.

## THE CENTRE FOR HEALTH LAW

Historically, the Law School has informally grouped health law teaching in the Centre for Health Law within the Department of Academic Legal Studies (DALS). The Centre, in its early years, received outside financial sponsorship from a USA-based clinical risk management company, MMI, and this helped develop professional linkages with a number of international and UK-based health bodies. The Director of the Centre of Health Law, John Tingle, Reader in Health Law, has also attracted outside income to the university through his editorship of *Health Care Risk Report*, a national monthly journal looking at the interface between risk, health law and health, published by Lexis Nexis. John was editor for seven years until May 2001. He continues his long association with the journal as a regular contributing author and a member of the editorial board.

The Centre has been very active for a number of years in short course development and consultancy for health care organizations and health care professionals. Courses have been held on a variety of issues including:

(a) Legal Aspects of Nursing;
(b) Legal Aspects of Clinical Risk Management and Clinical Governance;
(c) Expanded Role of the Nurse;
(d) Clinical Guidelines and the Law;
(e) Handling Patient Complaints;
(f) AIDS and the Law.

## THE ANNUAL CENTRE FOR HEALTH LAW CONFERENCE

Every year the Centre hosts a major conference on a topical area of health law with a national sponsor:

Previous events include:

(a) Health Law and Human Rights (2000), sponsor, Bird and Bird solicitors, London;
(b) Health Law and Our Children (2001), the Medical Protection Society;
(c) Health Law and the Regulation of Health Care Quality (2002), the Medical Protection Society.

In 2003 the University hosted the SLSA (Socio-Legal Studies Association) Annual Conference and the health law team organized the health law streams for this event. A number of health law papers presented at the conference were subsequently developed into book chapters and will appear in *Contemporary Issues in Healthcare Law and Ethics*, edited by Austen Garwood-Gowers, Kay Wheat and John Tingle, (Elsevier, Oxford). The book is in press and will be published in the first half of 2005.

## THE STRASBOURG SUMMER SCHOOL PROGRAMME

Every year in June, the Centre for Health Law runs a one-week programme for Law School students in Strasbourg, France. The programme introduces students to key

European institutions such as the Council of Europe, European Court of Human Rights and the European Parliament. The Summer School takes a particular look at the work of the Council of Europe on bio-ethics and human rights. Our students are joined by law students from Loyola University, Chicago, and they have joint visits, some joint classes and social activities. Interest in the programme from our students continues to grow and in 2005 numbers have doubled, with about 55 students joining the programme.

The Law School is expanding its summer school programmes and for the first time in 2005 there will be a programme based in Tallinn, Estonia. The Tallinn programme focuses on EU and international trade-related issues. This programme has attracted over 20 students for 2005 and looks set to follow the success of the Strasbourg programme.

## RESEARCH

Health law research members include: Austen Garwood-Gowers; Michael Gunn; Liz Rodgers; John Tingle and Kay Wheat.

Research areas include:

(a) Capacity and the law as it relates to people with learning disabilities and mental health problems generally, (this work includes that of Professor Gunn with his colleagues at the University of Cambridge, Department of Psychiatry (Professor Tony Holland and Dr Isabel Clare));
(b) Ethical and legal aspects of medical research and transplantation (in particular the work of Dr Garwood-Gowers and Ms Wheat);
(c) Children and the law;
(d) Medical and Nursing Negligence and Clinical Risk Management, in particular the work of John Tingle.

Health law has a bright future at Nottingham Law School. Significant numbers of academic staff and very large numbers of students clearly enjoy the subject and this has led to major publications and teaching programmes being developed.

JOHN TINGLE*

* BA Law Hons, Cert Ed (Dist), MEd, Barrister, Reader in Health Law, Director of the Centre for Health Law, Nottingham Law School.

# TABLE OF CASES

# TABLE OF STATUTES

# TABLE OF STATUTORY INSTRUMENTS

# TABLE OF EU LEGISLATION

# NOTTINGHAM LAW JOURNAL

VOL 14(2) 2005


## CONTENTS

NOTTINGHAM LAW JOURNAL

VOL 14(2) 2005

# EDITORIAL

There is apparently a movement, within the world of the commercial print media, to produce the occasional "student edition" involving a replacement and usually much younger, editorial team. The journal hardly qualifies for such schemes but for once the assistant editor has captured the editorial page, so allowing for proper recognition by the *Nottingham Law Journal* of a matter that would otherwise be apparent only to those prepared to scrutinise the first page in some detail. Adrian Walters, under whose editorship the journal has been flourishing since 2003, has been awarded a well-deserved chair in Corporate and Insolvency Law.

Continuing the theme of celebration of home-grown achievement, it is pleasing to see that this edition particularly reflects the breadth of the Law School and its activities with contributions from Andrew Alonzi and Professor Mark Mildred representing the graduate and professional divisions of Professional Legal Studies; from Academic Legal Studies in the shape of Simon Boyes (in two submissions both representing the Law School's commitment to sports law); from Angela Donaldson, one of our colleagues in the university library; as well as from Lisa Rodgers, a practitioner, and from a selection of other varied and eminent contributors from outside the university, writing on a wide spectrum of topics. Mirko Bagaric and James McConvill pose relevant questions about the interpretation of international treaties in the highly current field of asylum from the Australian perspective; Jane Johnson and Geraldine Hammersley present unusual qualitative data derived from the rare use of US-style contingency fees in a UK context and Gary Scanlan debates problems in the intellectual property of designs. Long may the journal continue to flourish so abundantly.

JANE CHING

NOTTINGHAM LAW JOURNAL

VOL 14(2) 2005

# ARTICLES

## REFUGEE LAW: THE IRRELEVANCE OF THE FRAMERS' INTENTIONS

*MIRKO BAGARIC* and JAMES McCONVILL***

### INTRODUCTION

The Convention relating to the Status of Refugees 1951 ("the Convention") is over fifty years old. It is the most comprehensive legally binding international standard for the treatment of refugees.[1] The Convention governs the rights of refugees and the obligations of ratifying State parties towards refugees. The key aspect of the Convention is article 1A(2), which sets out the Convention definition of a refugee. It provides that a refugee is a person who:

> Owing to a well-founded fear of being persecuted for reasons of race, religion, nationality, membership of a particular social group or political opinion, is outside the country of his or her nationality and is unable or, owing to such fear, is unwilling to avail himself of the protection of that country; or who, not having a nationality and being outside the country of his former habitual residence, is unable or, owing to such fear, is unwilling to return to it.[2]

At the international law level the definition has remained (effectively)[3] unchanged during this period. However, there has been a considerable amount of uncertainty at the domestic level concerning the precise meaning that should be given to important aspects of the definition, such as "particular social group" and "persecution".

Given that the Convention is the principal international instrument dealing with the rights of refugees (since it was ratified by Denmark in 1952, 140 states have acceded to the Convention)[4] and the importance of the interests and indeterminate nature of several aspects of the definition of refugee, the interpretive approach adopted in relation to the Convention is of considerable importance.

[1] P Weis, *The Refugee Convention*, 1951: *The Travaux Préparatoires Analysed* (1995).

[2] Convention relating to the Status of Refugees 1951, art 1A as amended by the Protocol relating to the Status of Refugees 1967, art 1.

[3] The only change is pursuant to the Protocol of 1967 which made no substantive changes to the definition. It merely removed a temporal and geographic limitation.

[4] DIMIA, *Interpreting the Refugees Convention – an Australian Contribution* (2002, Australian Government Printer).

In interpreting the definition, the Australian courts have at times placed emphasis on the intention of the framers of the Convention. In this article, we suggest that such an approach is misguided and consider several arguments against attributing relevance to the framers' intent.

The first, which we consider in the third section of the article, is the pragmatic obstacle that the concept of a group intention is incoherent. We argue that while there are conceptual difficulties associated with the concept of a group intention, these difficulties are not decisive; hence theoretically, it is possible to have regard to the intention of the framers of the Convention.

However, we argue that to do so would be unsound. There are two reasons for this. These are considered in the fourth part of the article. The first reason why the intention of the framers should be ignored is that such an approach would detract from the humanitarian underpinning of the Convention. The considerations that underpin forced people movement today differ greatly from those that forced people to flee their homelands over half a century ago. The concerns, issues and motives that weighed on the minds of the framers of the Convention are no longer relevant in today's international environment. The Convention was drafted against the background of Cold War politics and reflects the European experience of Nazi persecutions and Western political interests as they existed at the time: "it was not designed with today's mass refugee outflows or migratory movements in mind".[5] As is discussed below, in light of these sentiments, there have been recent calls that the Convention is no longer relevant.

Secondly, the interpretation that is given to a legal document should not be divorced from its source of authority. In the case of the Convention, the authority of the document stems from the decision of the Australian parliament to incorporate it as part of the domestic law of the country.

In the concluding remarks we make some observations concerning how the Convention should be interpreted to best deal with contemporary global humanitarian problems. Prior to turning to substantive issues, we first provide a brief overview of the background to the Convention, its relevance today and the recognition that the courts currently give to the framers' intention when engaged in the process of interpreting the Convention.

## BACKGROUND TO THE CONVENTION AND RELEVANCE OF THE FRAMERS' INTENTION

### *Historical Account*

The protection of refugees under international law commenced in the early 20th century when treaties were formulated to address refugee problems. Prior to this time, customary international law imposed an obligation on states to protect their own nationals only. This obligation did not extend to individuals from other nations who found themselves within the borders of a state. States had the discretion to accept immigrants whom they perceived would contribute to the economy or society in a positive way, and to expel refugees under the assumption that the right to do so was inherent in a state's sovereign powers.[6]

[5] A Millbank, *The Problem with the 1951 Refugee Convention* (2000, Australian Government Printer).

[6] Thomas Musgrave, "Refugees", in Sam Blay, Ryszard Piotrowicz and B Martin Tsamenyi (eds), *Public International Law: An Australian Perspective* (1997, OUP) 301.

Between 1919 and 1939, numerous violent conflicts and political problems in Europe and the Middle East led to the displacement of large numbers of people.[7] This enormous exodus clashed with the desire of states to control the immigrants entering their borders and led the international community to having to respond to the refugee issue. The League of Nations did so by formulating agreements to provide for refugee protection.[8] These instruments were *ad hoc* in nature, in that they related to *specific* refugee situations. They adopted a group or category approach, where the sufficient and necessary conditions to achieve refugee status were that someone was (a) outside their country of origin and (b) without the protection of the government of that state.[9] There was neither a general definition of refugee status, nor any standardised measure of international protection for refugees during this period.[10]

When masses of people were uprooted after World War II, it was perceived that the refugee problem was not a temporary one and that an instrument with a broader approach would more effectively address emerging refugee crises. Thus, the Convention relating to the Status of Refugees was adopted by a special United Nation Conference on 28 July 1951[11] and entered into force on 21 April 1954. It was drafted between 1948 and 1951 by a combination of United Nations organs, *ad hoc* committees and a conference of plenipotentiaries at which 26 states were represented.[12] The records of the negotiations (the *travaux préparatoires*) are recorded in various forms,[13] thereby making it tenable to gain insight into the intentions of the framers.[14] The fact that so many different parties and interest groups contributed to the drafting necessarily reduced the prospect that the definition would be based on a coherent overarching theory: this is certainly the way matters ultimately transpired.

Unlike earlier *ad hoc* instruments, the Convention purported to provide a *general* definition of who was to be considered a refugee and in addition provided a guarantee of non-refoulement, whereby refugees could not be returned to their country of origin if doing so would subject them to persecution within the meaning of the Convention.

Despite its universal overtones, the 1951 Convention was limited by the fact that it protected mainly Europeans fleeing after the war[15] and by the dateline contained within the definition. These restrictions were removed and the definition was expanded (formally, though not substantively), by the 1967 Protocol relating to the Status of Refugees.[16] Accession to the Protocol enabled states to apply the substantive provisions of the Convention to refugees as defined by the Convention, but without the dateline and geographic limitations. Hence, the Convention now applies to all persons who are refugees because of events occurring at *any time*, not merely those in Europe before 1951. The Protocol is an instrument independent of the Convention and accession to it is not limited to states which are parties to the Convention.

[7] *Ibid.*

[8] UNHCR, *The State Of The World's Refugees 2000: Fifty Years of Humanitarian Action* (2000, OUP) 15.

[9] Guy Goodwin-Gill, *The Refugee in International Law* (2nd ed, 1996, Clarendon Press) 4.

[10] Musgrave, above n 6, 302.

[11] Goodwin-Gill, above n 9.

[12] James Hathaway, *The Law of Refugee Status* (1991, Butterworths) 6. For a history of the process leading to the drafting of the convention, see Weis, above n 1, 1–4. Note the earlier international agreements entered into on behalf of refugees are referred to in art 1A(1) of the Convention.

[13] See, for example, Weis, above n 1.

[14] The concept of the framers' intentions is discussed further below.

[15] Goodwin-Gill, above n 9. Note that the definition included an optional geographical limitation that permitted states, on ratification, to limit their obligations to refugees from "events occurring within Europe" prior to the critical date: art 1B.

[16] Hathaway, above n 12, 10.

It is clear that the Convention was entered into mainly to assist European refugees, and to serve Western political and economic needs.[17] This is a point emphasised by Hathaway:

> The two main characteristics of the Convention refugee definition are its strategic conceptualisation and its Eurocentric focus. The strategic dimension of the definition comes from successful efforts of Western States to give priority in protection matters to persons whose flight was motivated by pro-Western political values. As anxious as the Soviets had been to exclude political émigrés from the scope of the Convention for fear of exposing their weak flank, so the more numerous and more powerful Western states were preoccupied to maximise the international visibility of that migration. In the result, it was agreed to restrict the scope of protection in much the same way as had been done in the post-World War II refugee instruments: only persons who feared "persecution" because of their civil or political status (me) would fall within the international protection mandate. This apparently neutral formulation facilitated the condemnation of Soviet bloc politics through international law in two ways. First, the persecution standard was a known quantity, having already been employed to embrace Soviet bloc dissidents in the immediate post-war years. Second, the precise formulation of the persecution standard of the persecution meant that refugee law could not readily be turned to the political advantage of the Soviet bloc. The refugee definition was carefully phrased to include only persons who have been disenfranchised by their state on the basis of race, religion, nationality, membership of a particular social group, or political opinion, matters in regard to which eastern bloc practice has historically been problematic. Western vulnerability in the area of respect for human rights, in contrast, centers more on the guarantee of socio-economic human rights, than on respect for civil and political rights. Unlike the victims of civil and political oppression, however, persons denied even such basic rights as food health care or deduction are excluded from the international refugee regime (unless that deprivation stems from civil or political status). By mandating protection for those whose (Western inspired) socio-economic rights are at risk, the Convention adopted an incomplete and politically partisan human rights rationale ... In addition to their desire for the refugee Convention to serve strategic political objectives, the majority of the States that drafted the Convention sought to create a rights-regime conducive to the redistribution of the post-war refugee burden from European shoulders. [18]

Thus the history of the Convention reflects the fact that the plight of refugees was in fact subordinated to the needs and wants of state parties drafting the Convention. The definition was created by self-interested nation states pre-occupied with Cold War politics.[19] It has been noted by Hathaway that:

> it remains tragically true that international human rights law – the intended means of permitting the world community to respond to wrongs committed by a country within its own territory – has not been permitted to evolve to a state of genuine efficacy.[20]

Given the historical backdrop and motivation to the Convention it is not surprising that some are starting to question its relevance in today's world.

In 1998 the Austrian Presidency of the EU suggested replacing the Convention with an EU asylum law "which meets today's requirements rather than those of a geopolitically outdated situation". In the same year the General Secretary of Germany's Liberal Party called in effect for default from the Convention on the grounds that it was "an invitation to abuse and to unrestricted and unregulated migration". In April this year the UK Home Secretary, Jack Straw, criticised the

[17] Hathaway, above n 12, 6.

[18] Hathaway, above n 12, 7–8 (footnotes omitted).

[19] *Ibid* 232–233.

[20] *Ibid* v.

Convention as "too broad for conditions in the 21st century", and as "no longer an adequate guide to policy in the age of mass air travel and economic migration". Conservative Party leader William Hague described the asylum system as "near collapse in today's utterly different world".[21]

In light of such criticism the manner in which the Convention is interpreted is particularly important. Prior to considering the approach to interpretation that should be adopted, we consider, in the third section, whether it makes sense to contemplate an approach based on the framers' intention. Prior to doing so, we briefly consider the current approach to interpreting the Convention.

*Interpretation of Convention*

The interpretation of international law documents is generally governed by the Vienna Convention on the Law of Treaties. Article 31(1) provides that: "A treaty shall be interpreted in good faith in accordance with the ordinary meaning to be given to the terms of the treaty in their context and in the light of its objects and purpose".

Article 31(4) provides that "a special meaning shall be given to a term if it is established that the parties so intended".

In terms of the significance of article 31 in a domestic law context, the High Court of Australia has held that it forms the basis for treaty interpretation unless the legislature has expressed an intention to the contrary. In *Applicant A and Anor v MIEA and Anor*, Dawson J said:

> Deciding that question [the interpretation of the Refugees Convention] involves the construction of a domestic statute which incorporates a definition found in an international treaty. Such a provision, whether it is a definition or otherwise, should ordinarily be construed in accordance with the meaning to be attributed to the treaty provision in international law. By transposing the provision of the treaty, the legislature discloses the *prima facie intention that it have the same meaning in the statute as it does in the treaty.* Absent a contrary intention, and there is none in this case, such a statutory provision is to be construed according to the method applicable to the construction of the corresponding words in the treaty. ... The general rule of interpretation of treaty provisions appears in article 31 of the Vienna Convention on the Law of Treaties (the Vienna Convention) ... Under that rule, the starting point must be the text of the treaty. Of course, the text of the treaty is often couched in fairly general terms due to differences in language and legal conceptions among those to whom it is to be addressed and as part of an attempt to reach agreement among diverse nations. Accordingly, technical principles of common law construction are to be disregarded in construing the text. Article 31 plainly precludes the adoption of a literal construction which would defeat the object or purpose of the treaty and would be inconsistent with the context in which the words being construed appear. To say as much is, perhaps, to state no more than the accepted canon of construction that an instrument is to be construed as a whole and that words are not to be divorced from their context or construed in a manner which would defeat the character of the instrument.[22]

In the same case, McHugh J stated that a holistic approach should be taken to treaty interpretation.[23] This means that "primacy is to be given to the written text of

[21] Millbank, above n 5.

[22] (1997) 190 CLR 225, 239–40. Murphy J in *The Commonwealth v Tasmania (Tasmanian Dams Case)* stated that the relevant international convention (in that case the UNESCO Convention for the Protection of the World Cultural and National Heritage) should be interpreted in a manner that gives "primacy to the ordinary meaning of its terms in their context and in light of its objects and purpose": (1983) 158 CLR 1, 177.

[23] *Applicant A and Anor v MIEA and Anor* (1997) 190 CLR 225, 253–254, see also 231 (Brennan CJ).

the Convention but the context, object and purpose of the treaty must also be considered".[24] In a similar vein, Brennan CJ stated:

> In interpreting a treaty, it is erroneous to adopt a rigid priority in the application of interpretive rules. The political processes by which a treaty is negotiated to a conclusion preclude such an approach. Rather, for the reasons given by McHugh J, it is necessary to adopt an holistic but ordered approach. The holistic approach to interpretation may require a consideration of both the text and the object and purpose of the treaty in order to ascertain its true meaning. Although the text of a treaty may itself reveal its object and purpose or at least assist in ascertaining its object and purpose, assistance may also be obtained from extrinsic sources. The form in which a treaty is drafted, the subject to which it relates, the mischief that it addresses, the history of its negotiation and comparison with earlier or amending instruments relating to the same subject may warrant consideration in arriving at the true interpretation of its text.[25]

In essence, the principal guide to the interpretation of international treaties, including the Convention, is the text of the document. However, this is to be complemented by considerations pertaining to the objects and purposes of the treaty. However, given the primacy of domestic law (including principles of statutory interpretation) over international law documents, the legislature is free to depart from general principles of treaty interpretation and to change the meaning of a treaty when it interprets it into domestic law.

To this end, the principal domestic law governing statutory interpretation in Australia is the Acts Interpretation Act 1901 (Cth). Pursuant to section 15AA, legislation should be interpreted in a manner which promotes the purpose or object underlying the Act.[26] In this regard the emphasis is on the domestic law, and hence there is less scope for giving effect to the intentions of the framers of an international legal instrument incorporated into domestic law. However, this does not totally forestall inquiry into the objects of the international instrument. As is noted by Dawson J above, unless a contrary intention appears in the statute, it should be given the same meaning as in the treaty.[27] Moreover, in *Minister for Foreign Affairs and Trade v Magno*,[28] Gummow J emphasised that where an international agreement has been incorporated into Australian domestic law its interpretation is governed by articles 31 and 32 of the Vienna Convention and not by sections 15AA and 15AB of the Acts Interpretation Act.[29] Thus, the objects of the international instrument can only be rendered irrelevant if this is made clear, by express words or necessary

[24] *Ibid* 254, see also 277 (Gummow J), 292–6 (Kirby J). For further discussion of principles of interpretation of the Convention see also *MIMA v Haji Ibrahim* (2000) 204 CLR 1 and *MIMA v Khawar* (2002) 187 ALR 574. In *MIMA v Savvin* (2000) 171 ALR 483, 502 (Katz J) and 487 (Drummond J) it was held that the Vienna Convention did not apply to the interpretation of the Convention because the Convention (and Protocol) was concluded by Australia after the entry into force of the Vienna Convention. However, it was held that even though the Vienna Convention is not applicable in the construction of the Convention, still the Vienna Convention "constitutes an authoritative statement of customary international law" (at 502). See also *Hellman v MIMA* (2000) 175 ALR 149, 154 where it was held that the Convention is to be interpreted in accordance with the requirements of the Vienna Convention.

[25] (1997) 190 CLR 225, 231.

[26] For a detailed examination of this clause, see D C Pearce and R S Geddes, *Statutory Interpretation in Australia* (1996, Butterworths), Ch 2.

[27] *Applicant A and Anor v MIEA and Anor* (1997) 190 CLR 225, 230–231; also 239 (Brennan CJ).

[28] (1992) 112 ALR 529. See also cases noted in Pearce and Geddes, above n 26, 45–46.

[29] Additionally, the Acts Interpretation Act 1901 (Cth) s 15AB(2)(d) expressly permits recourse to "any treaty or other international instrument that is referred to in the Act" to confirm the ordinary meaning or where there is an ambiguity. As Gleeson CJ recently pointed out in *Plaintiff S157/2002 v Commonwealth of Australia* [2003] HCA 2, at [27], the Migration Act is "part of a wider ... pattern of legislation which, ... affects fundamental human rights and involves Australia's international obligations".

implication in the domestic legislation dealing with the instrument. This has not been done in relation to the Convention and the courts have therefore on occasions had regard to the intentions of the founders in its interpretation.

A broad statement of the use courts make of the framers' intentions in interpreting the Convention is made by Goodwin-Gill, who notes:

> For better of worse, refugee status decision makers (and commentators . . .) make frequent use of the *travaux préparatoires* to the 1951 Convention. Many key terms are vague, undefined and open to interpretation, but the results of inquiry into the background, as the present analysis shows, can be rather mixed. On the one hand, clear statements of drafting intentions are rare, yet on the other hand, the debates in the General Assembly, the Third Committee, the Economic and Social Council and, to a lesser extent, at the 1951 Conference itself, provide a fascinating insight into the politics of a highly sensitive and emotive issue. If some sentiments and statements seem frozen in time, others show the continuity of concern and, perhaps too rarely, confirmation of a pervasive humanitarianism.[30]

The Australian courts have on numerous occasions referred to the intention of the framers as a guide to the interpreting the Convention. When they have done so, they have not engaged in a detailed historical analysis of the documents that recorded the deliberations resulting in the Convention. By and large the reference to the framers' intent has been to broad ideals and concerns which the court has *assumed* were foremost in the minds of the framers at the time the Convention was drafted.[31]

However, in a small amount of cases the courts have gone further and placed considerable emphasis on the framers' intent. For example, in *MMM v Minister for Immigration and Multicultural Affairs* in the context of considering whether the criminalisation of homosexuality amounted to "persecution", the Federal Court of Australia stated:

> With respect, [a formal interpretation of persecution] . . . ought not be imputed to the framers of the Convention. There is, in ordinary language and common sense, a clear and cogent distinction between a mere infringement of an internationally recognised human right and persecution. It is not, in my opinion, sensible to ignore matters of degree, nor were the framers of the Convention likely to have done so. Many countries have some laws that are more honoured in the breach rather than the observance. Commonly these deal with matters seen by some as concerning only questions of private morality but by others as concerning important questions of standards legitimately the subject of public laws. As such they often raise questions of "human rights". Failure to enforce contentious laws is a common social lubricant. The matter of the criminal law and homosexual acts itself furnishes an example: in countries whose tradition is Western civilisation, private consensual homosexual acts were, until quite recently, generally penalised. That is no longer so, but liberalisation (even for heterosexual fornication – see *Z v Minister for Immigration and Multicultural Affairs* (unreported, Federal Court of Australia, Katz J, 11 December 1998) has not been universal. The framers of the Convention were concerned with persecution of a kind which morally obliged civilised States to receive refugees, regardless of other restrictions those States might place on immigration. The fourth preamble to the Convention recognised that "the grant of asylum may place unduly heavy

[30] Goodwin-Gill above n 9, 368

[31] For example, see *Minister for Immigration and Multicultural Affairs v Ibrahim* (2000) 204 CLR 1, 17–8, citing extracts from the House of Lords' decision in *Adan v Secretary of State for the Home Department* [1999] 1 AC 293, where the House of Lords looked to the framers' intention in interpreting the definition. In this regard, see also *Minister for Immigration and Multicultural Affairs v Abdi* (1999) 162 ALR 105, 113-4; *Qaraninamu v Minister for Immigration and Multicultural Affairs* [2000] FCA 1611 (13 November 2000), para 13; *Rukhiyah Farrah Mohammed v Minister For Immigration and Multicultural Affairs* [1998] 1077 FCA (3 September1998). Other broad references to framers' intent are made, for example, in *Minister for Immigration and Multicultural Affairs v Khawar* (2002) 210 CLR 1, 15–7 (McHugh and Gummow JJ); *Nagaratnam Prathapan v Minister for Immigration and Multicultural Affairs* [1997] 937 FCA (5 September 1997); *Mohamed Dahir Mohamed v Minister for Immigration and Multicultural Affairs* [1998] 485 FCA (11 May 1998); *Sarrazola v Minister for Immigration and Multicultural Affairs (No 3)* [2000] FCA 919 (23 August 2000), [34].

> burdens on certain countries". Merely to be legally stigmatised because the expression of one's (legitimate) sexual desires is subjected to the theoretical possibility of a heavy penalty, without proof of a real chance of more substantial harm, is hardly likely to have been of great concern to States dealing with the consequences of the murderous excesses of Nazi Germany or Stalinist Russia, and with being obliged to exclude from access to refugee status persons suffering from natural disasters and other sources of profound misery.[32]

In *Chan*, the High Court of Australia attempted to undertake a detailed analysis of the documents culminating in the Convention but was not assisted in this process. McHugh J in *Chan* sought to elucidate the meaning of "well-founded fear":

> The examination of *travaux préparatoires* to construe a treaty is a legitimate and perhaps a necessary tool for construing such a document. "It may now be regarded as a settled principle of interpretation of treaties that tribunals, international and national, will have recourse, in order to elucidate the intention of the parties, to the records of the negotiations preceding the conclusion of the treaty, the minutes of the conference which adopted the treaty, its successive drafts, and so on": Lauterpacht, *International Law*, vol 1 (1970), at p 363. Unfortunately, the preparatory materials are of limited assistance in interpreting the present Convention and Protocol.[33]

Thus, the courts have tried to discern the specific intention of the framers as regards the meaning of many of the broad terms used in the Convention, but without success. Despite the fact that the *precise* intent of the framers has had a limited effect on the interpretation of the definition of a refugee, there is no question that the contemporary political context that inspired the Convention has contributed to the narrowness with which the definition of refugee has been interpreted. For example, in *MIMA v Haji Ibrahim,* Gummow J considered the historical background to the Convention:

> The drafting of the Convention reflected the concerns of the Cold War. The participating States did not include the Soviet Union and its allies, which had objected to a notion of comprehensive protection for refugees. The States which did participate nevertheless had no commitment to basing the Convention in the international promotion of human rights.[34]

His Honour continued:

> It is generally accepted that the Convention definition, based on individual persecution, limits the humanitarian scope of the Convention. The definition does not encompass those

[32] (1998) 170 ALR 411, 415.

[33] *Chan v Minister for Immigration and Ethnic Affairs* (1989) 169 CLR 379, 428. His Honour further noted: "The definition of 'refugee' in the Geneva Convention has been traced to the International Refugee Organization ('IRO') Constitution adopted by the General Assembly of the United Nations in 1946: Cox, *op cit*, at p 338. The IRO Constitution stated that no refugee with 'valid objections' should be compelled to return to his or her country of origin. One of the 'valid objections' was persecution or 'fear based on reasonable grounds of persecution because of race, religion, nationality or political opinion . . .'. The IRO Manual declared that 'reasonable grounds' were to be understood as meaning that the applicant has given 'a plausible and coherent account of why he fears persecution'. Subsequently, the Economic and Social Council of the United Nations in 1949 appointed an *Ad Hoc* Committee on Statelessness and Related Problems. The Committee drafted a provisional Convention in 1950 and this led to the Convention that was done at Geneva in July 1951. In its Final Report to the Council, the Committee declared that it had drafted the Convention to afford at least as much protection to refugees as had been provided by previous agreements. This declaration might suggest that a 'well-founded' fear under the Convention is established, having regard to the IRO Manual interpretation, if the applicant gives a plausible and coherent account of why she fears persecution. But as Cox, *op cit*, has pointed out (p 351) the IRO approach was dictated by its inability to form an independent and objective view about conditions prevailing in the country of origin and the state parties to the Convention and Protocol will frequently have detailed knowledge of conditions in the country of the applicant's nationality. It is unlikely, therefore, that a state party was expected to grant refugee status to a person whose account, although plausible and coherent, was inconsistent with the state's understanding of conditions in his or her country of nationality. The practice of the IRO under its constitution, therefore, is no guide to the meaning of 'well-founded fear' in the Convention and Protocol definitions notwithstanding the declaration of the *Ad Hoc* Committee in its Final Report (at 428)".

[34] *MIMA v Haji Ibrahim* (2000) 204 CLR 1, 47.

> fleeing generalised violence or internal turmoil and mass movements of persons fleeing civil war or other armed conflicts, military occupation, natural disasters and bad economic conditions are outside the Convention.[35]

In *Applicant A*, McHugh J referred to the drafting history of the "particular social group" ground in seeking to identify the sort of group that the category was intended to cover:

> It seems likely that the category of "particular social group" was at least intended to cover those groups persecuted because of "the 'restructuring' of society then being undertaken in the socialist States and the special attention reserved for landowners, capitalist class members, independent business people, the middle class and their families". In *Bastanipour v INS* (1992) 980 F 2d 1129 at 1132), Posner J thought that the kulaks (affluent Russian peasants) who had been persecuted by Stalin were the sort of group intended to be covered by the term "particular social group". All the foregoing groups are disparate in character. But what distinguishes their members from other persons in their country is a common attribute and a societal perception that they stand apart. Persecution, of course, reinforces the perception that they are "a particular social group" in their country.[36]

Recently, the Australian legislature has been more prepared to impose its will on key aspects of the meaning of the Convention definition. Pursuant to relatively recent changes to the Migration Act 1958 (Cth), the term "persecution" in article 1A(2) is now qualified by section 91R. Section 91R(1) provides that for the purposes of the Act and regulations, article 1A(2) does not apply in relation to persecution for one or more of the Convention reasons unless:

> (a) that reason is the essential and significant reason, or those reasons are the essential and significant reasons, for the persecution;
> (b) the persecution involves serious harm to the person; and
> (c) the persecution involves systematic and discriminatory conduct.

Section 91R(1)(b) of the Act provides that:

> For the purposes of the application of this Act and the regulations to a particular person, Article 1A(2) of the Refugees Convention as amended by the Refugees Protocol does not apply in relation to persecution for one or more of the reasons mentioned in that Article unless:
> ... (b) The persecution involves serious harm to the person ...

Subsection (2) sets out a non-exhaustive list of the type and level of harm that can be classified as serious. The following examples of "serious harm" are provided:

> (a) a threat to the person's life or liberty;
> (b) significant physical harassment of the person;
> (c) significant physical ill treatment of the person;
> (d) significant economic hardship that threatens the person's capacity to subsist;
> (e) denial of access to basic services, where the denial threatens the person's capacity to subsist; and
> (f) denial of capacity to earn a livelihood of any kind, where the denial threatens the person's capacity to subsist.

[35] *Ibid* 49.

[36] *Applicant A and Anor v MIEA and Anor* (1997) 190 CLR 225, 265–266.

This definition of persecution reflects many of the themes in the case law,[37] which, at times, has been developed with explicit regard to the framers' intention. However, it can also be argued that some aspects of the statutory definition (especially the meaning of serious harm) in fact move away from the pre-existing case law.[38] Hence, the legislature is displaying a greater preparedness to stamp its mark on key aspects of the definition of a refugee. In our view, this approach is commendable. But it does not go far enough. The legislature should make it clear that in interpreting the Convention the intentions of the framers are irrelevant. This is discussed further in the fourth section of this article. Before considering whether regard *should* be had to the framers' intention, we consider the threshold argument of whether regard *can* be had to such intentions.

## THE CONCEPT OF GROUP INTENTION: IS IT COHERENT?

The strongest pragmatic objection to the ascription of weight to the framers' intention is that the concept of a group intention is incoherent. An intention is a state of mind: a personalised state of mind. It is wholly artificial and contrived, so the argument runs, to speak of a group (in this case, the collection of people who contributed to the drafting of the Convention) having an intention.

### *Analogy with Constitutional Law*

This is an issue that has aroused some attention in the context of the approach to statutory interpretation generally. It has generated most interest and commentary in the context of the interpretation of one particular type of statute: the Australian Constitution. In this regard obvious parallels can be drawn between the viability of ascertaining the intention of the founders of the Constitution and that of the framers of the Convention.[39]

### *Rejection of Group Intention*

The issue of framers' intention has arisen most pointedly in the context of the "originalism" debate. Originalism is the theory that a constitution should be interpreted to give effect to the framers' intentions.[40] The originalism debate has generated a mountain of literature in the United States over the past few decades – in the process inciting an enormous amount of passion. Originalists have labelled alternative methods of constitutional interpretation as "heresy",[41] while their opponents have claimed that originalism should be stored away "with phlogistonism and

[37] M Bagaric and P Dimopoulos, "Prosecution and Persecution: towards a workable dichotomy in refugee law – Discrimination as the touchstone of the nexus between serious harm and the Convention grounds" (2004) 32 *International Journal of the Sociology of Law* 310.

[38] M Bagaric and P Dimopoulos, "The Shifting Meaning of Persecution in Australian Refugee Law: how much must one suffer to be deserving of asylum?" (2003) 15 *Bond Law Review* 284.

[39] For a detailed discussion concerning the issue of group intent in the context of the originalism debate, see M Bagaric, "Originalism: Why Some Things Should Never Change – or at Least not Change too Quickly"(2000) 19 *University of Tasmania Law Review* 173.

[40] When the originalist methodology is applied in the context of normal statutes, it has been labelled "intentionalism": see D Lyons, "Original Intent and Legal Interpretation" (1999) 24 *Australian Journal of Legal Philosophy* 1.

[41] R Bork, *The Tempting of America: The Political Seduction of the Law* (1990, Free Press) 159.

[the] bogeyman".[42] Many have been prepared firmly to dismiss originalism: "within the scholarly community originalism is in disrepute and . . . anyone seeking to defend the theory must do so with trepidation".[43]

Michael Kirby labels the opposing approach to originalism as the "living force"[44] interpretive method and neatly encapsulates the tension between the methodologies:

> [One view is] that a constitution is anti-evolutionary. The other is that it must be evolutionary. The one that it is necessary to anchor a constitutional text and its meaning in the ascertainable fact of the intentions of the drafters in an earlier century. The other, that this involves a primitive form of ancestor worship, inappropriate to constitutional interpretation in a modern state. The one believing that legitimacy can be found, and found only, in legal history. The other believing that a constitution is a living tree which continues to grow and to provide shelter in new circumstances to the people living under its protection.[45]

Until recently the originalism issue remained virtually dormant in Australia. However, since the High Court ended its self-imposed ban on referring to the Convention Debates[46] about a decade ago,[47] the landscape has quickly changed.[48] One of the main criticisms of originalism is that the theory is pragmatically untenable because the concept at its core – the framers' intent – is either unintelligible or unascertainable.

The weight of contemporary academic opinion is that the concept of a group intention is a nonsense. Ronald Dworkin argues that the difficulties involved in ascertaining the framers' intentions are so overwhelming that in reality "there is no intention, in such a body, even in principle, only one waiting to be invented".[49] The criticisms of the concept of a group intention are thought to be so compelling that McHugh J in *Eastman v The Queen* stated that "*no doubt* the notion of constitutional intent, like legislative intent, is fictitious".[50]

[42] R Dworkin, "Bork's Jurisprudence" (1990) 57 *The University of Chicago Law Review* 657, 674.

[43] E Maltz, "The Appeal of Originalism" (1987) 4 *Utah Law Review* 773. For an overview of the state of the debate in the context of the United States Constitution, where it is perceived that non-originalists have the upper hand, see J Goldsworthy, "Originalism in Constitutional Interpretation" (1997) 25 *Federal Law Review* 1, 21–25.

[44] M Kirby, "Constitutional Interpretation and Original Intent: A Form of Ancestor Worship?" (2000) 24 *Melbourne University Law Review* 1, 11.

[45] See M Kirby, "Constitutional Interpretation and Original Intent: A Form of Ancestor Worship?" (2000) 24 *Melbourne University Law Review* 1, 6.

[46] A complete record of the Convention Debates has always been available. More recently a fully indexed version has been published: G Craven (ed), The Convention Debates 1891–1898: Commentaries, Indices and Guide (1986, Australian Government Printer, 6 vols).

[47] The High Court of Australia has always, theoretically, been free to use any extrinsic aids it wished to assist it in interpreting the constitution: see H Burmester, "The Convention Debates and the Interpretation of the Constitution" in Craven, above n 45, 26. However, in *Municipal Council of Sydney v The Commonwealth* (1904) 1 CLR 108, the High Court imposed a ban which lasted until *Cole v Whitfield* (1988) 165 CLR 360. The curious feature of the court's resolve not to use the *Convention Debates* is that, when it found history relevant, it had no hesitation in referring to Quick and Garran (*The Annotated Constitution of the Australian Commonwealth* (1901, Australian Government Printer), which in effect summarises the intentions of the drafters. Not surprisingly, this approach has been persuasively criticised: if history is relevant it "should be as complete as practicable, not the kind of half hearted history which the court has indulged in over the years", Michael Coper, "The Place of History in Constitutional Interpretation" in Craven (ed), above n 45, 1015. See also *R v Pearson; ex parte Sipka* (1983) 45 ALR 1, 14–15 (Murphy J).

[48] For example, see G Craven, "Original Intent and the Australian Constitution – Coming Soon to a Court near You?" (1990) 1 *Public Law Review* 166; D Tucker, "Textualism: An Australian Evaluation of the Debate between Professor Ronald Dworkin and Justice Antonin Scalia"(1999) 21 *Sydney Law Review* 567; S Donaghue, "The Clamour of Silent Constitutional Principles" (1996) 24 *Federal Law Review* 133; Goldsworthy, above n 42, Kirby, above n 44; J Kirk, "Constitutional Interpretation and a Theory of Evolutionary Originalism" (1997) 27 *Federal Law Review*; D Meager, "Guided by Voices? Constitutional Interpretation on the Gleeson Court" (2002) 7 *Deakin Law Review* 261; D Meagher, "New Day Rising? Non-Originalism, Justice Kirby and Section 80 of the Constitution" (2002) 24 *Sydney Law Review* 141; J Goldsworthy, "Interpreting the Constitution in its Second Century" (2002) 24 *Melbourne University Law Review* 677.

[49] R Dworkin, "The Forum of Principle" (1981) 56 *New York University Law Review* 468.

[50] *Eastman v The Queen* (2002) 203 CLR 1, 46.

The strongest argument against the concept of a group intention is that in any group people will often have different intentions: there will be a conflict of intentions. This problem can be so acute that in some cases a law that is passed or a motion that is carried can, in fact, it is claimed, lack a majority of "intention votes".[51] Thus, for example, it can be argued that, if an aspect of the refugee definition had been included as a political compromise, then adherence to the framers' intent would entail ignoring this part of the definition.

*A Framework for Giving Meaning to Group Intention*

To deal with this, the first point that needs to be emphasised is that, irrespective of the (admittedly almost infinite) number of contingencies which could have resulted in a provision being passed, an indispensable feature of any (valid) legislative provision or motion carried by any other group is that it was in fact passed. The legislature, however constituted, and for whatever reason, ultimately decided to approve the relevant standard.

a) *Means and ends.* This seemingly innocuous observation serves as a partial counter to the criticism that some provisions may have been passed without the requisite number of intention votes. The claim that provisions are invalid if passed primarily as a result of a political compromise (or for some other reason did not enjoy the express approval of a majority of delegates) rests on the flawed assumption that decisions made, or acts performed, which are not ends in themselves, but rather are means to promote some further ends, are not intended. Most of the decisions we make and actions we perform are a means to further particular ends. The reason we drive our car each morning is not because of the intrinsic joy of driving, but rather in order to get to work. It is nonsensical in such circumstances to assert that we have no intention to drive. In all other respects, this is a fact accepted by the law. For example, it would be farcical for an accused charged with murder to assert that he or she did not "intend" to kill the victim who was shot in the head because he or she "only" did it as a means to rob the victim. Hence, groups that pass a proposal are properly regarded as being in favour of it, even though they may be indifferent or even hostile to it.

b) *Intentions and desires.* The claim that a person or group can cause a certain outcome (such as enacting a law) without properly intending that result also stems from a failure to distinguish intentions from desires.

Desires and intentions are often used interchangeably. For example, when a person commits murder, the mental state of the agent immediately preceding this act is often characterised as either an intention to kill or a desire to kill. But the concepts are in fact different. Desires are representations of how the world is to be; they are our wants, the states that move us to act. Desires are often contrasted with beliefs. Beliefs are copies or replicas of the way we believe the world to be. On their own, beliefs can never provide a source of motivation: "they are perfectly inert, and can never either prevent or produce any action".[52] It is only our desires that can motivate us. Beliefs are mere replicas of the way we believe the world to be. We can assess beliefs for truth and falsehood: a true belief being one which is a copy of the way the world actually is. In order for an action to occur we need a desire that prompts us to affect a certain change in the world and a belief informing us how this change can be achieved.

[51] P Brest, "The Misconceived Quest For Original Understanding" (1980) 60 *Boston University Law Review* 204, 213.

[52] D Hume, *A Treatise of Human Nature* (1738, revised edition 1978, OUP) 458. For a discussion of the relevance of the belief/desire distinction in moral philosophy, see M Bagaric, "Internalism and the Part-time Moralist: an essay about the objectivity of moral judgments" (2001) 2 *Journal of Consciousness and Emotion* 255.

An intention is a special type of desire: it is a mental predicate;[53] a *commitment* to achieving a state of affairs or action.[54] Put differently, it is the result or outcome an agent wants to achieve in order to satisfy all of his or her desires which are relevant to the particular matter, and which has resulted from the weighing, conflating and prioritising of all of the agent's relevant desires. Even though the agent ultimately kills the victim, there may be *countless* desires which are relevant to this decision, some of which may be conflicting. For example, a potential murderer may be moved by the desire to obtain revenge or money, while at the same time also desiring not to harm others and to avoid being punished. Despite the conflicting nature of these desires, there is no difficulty in accepting that they can co-exist simultaneously. Where such desires do exist, whether or not the agent ultimately kills the victim depends on the number and intensity of the relevant desires. However, in the end, despite the large number of desires that may exist, there will be only one intention in respect to any particular decision that must be made. The agent will either form an intention to kill or not to kill.

It follows that intentions cannot conflict (they cannot be inconsistent): it is incoherent to assert that an agent can be both committed and uncommitted to the same matter – it is nonsensical to claim that an agent both intends to kill and not to kill the victim. There is no such absurdity in the assertion that an agent both desires to eat a chocolate bar and desires not to eat it. In such a case, the agent is in a "pre-intention" state of mind and going through the mental exercise of weighing and evaluating the relevant desires.

c) *Analogy between individual and group intention.* The observation that individuals often have a number of different (and often conflicting) desires in relation to any particular matter, and yet are still capable of forming an intention concerning the matter can be used as a basis for explaining the concept of group intention. The key to ascribing a group intention is to "personify" the group and treat all of the different desires of each member of the group relating to the relevant decision in the same way as if they were found in an individual. In a group there will no doubt normally be a larger range of desires in respect to any particular matter and there will invariably be more conflicting desires, but this is only a difference in degree, not nature. As we saw, in order for an intention to be ascribed, there is logically no limit to the number and type of desires which can relate to the subject matter of the intention.

More fully, the concept of a group intention can be rationalised in the following way. The text of the relevant document (in this case the Convention) identifies the subject matter, and the desires which are relevant are all those which relate to the relevant subject matter. All of the intentions of the individual framers concerning the relevant subject matter are pooled together. Each of the framers' intentions is then accorded the same status as a normal desire within an individual. The only difference from the individual context is that the intention of each individual is no longer treated as an "intention" (*ie*, a commitment to achieving a certain outcome), because the intention of any particular individual is not necessarily decisive, but rather it is treated as a penultimate desire in an individual. The intentions of each of the framers are then evaluated for content and weight, and the group intention is then the outcome that an individual with all of the intentions (now penultimate desires) of the group would seek

[53] A Marmor, *Interpretation and Legal Theory* (1992, revised edition 1994, Clarendon Press) 162.

[54] For a discussion of the content of an intention and an overview of the philosophical debate in this area, see P R Cohen, J Morgan and M E Pollack, "Persistence, Intention, and Commitment" in P R Cohen, J Morgan and M E Pollack (eds), *Intentions in Communication* (1990, MIT Press) 32. Readers are also directed to a series of other papers in the book.

to achieve after weighing and conflating all of the intentions of the group. While ontologically it is obviously beyond the realms of human experience to ascertain with certainty the result of such a process, logically the analogy is valid.[55]

The above model can be used to counter the criticisms advanced concerning the concept of a group intention. There is no need for a majority of delegates expressly to support a provision for it to have the backing of an international consensus. It is not simply a matter of stacking up the numbers, otherwise it would be open for an accused to resist a murder charge by asserting that the intention to kill was absent because he or she had two desires against killing and only one (albeit strong) desire in favour of killing. Thus, for example if one delegate only wanted to include particular social group as a Convention ground and the remaining (say) ten delegates disagreed with the provision however allowed it to pass as a political compromise, then the sentiments of the ten delegates who were not in favour of the provision are equivalent to ten weak desires, which are trumped by the strong desire of the delegate who supports the law, thereby ultimately resulting in an intention to pass the provision. It is therefore flawed to claim that provisions are passed which lack a majority of intention votes. The intentions of those who voted against the ground are not relevant in a meaningful sense: they have been outweighed by the intention of the delegate in favour of the law.

Thus, in defending the concept of a group intention, we are not suggesting that ontologically such a concept exists. Rather, it is argued that (a group of) individuals can have similar or identical intentions in relation to a particular activity or subject matter. While group intention cannot be reduced to individual intentions, in a *meaningful* sense group intention is closely analogous to individual intention. By meaningful we mean that it is *logically* possible to extend the concept of an individual intention to a group setting.

It follows that the difficulties associated with group intention do not provide an overwhelming reason to dismiss out of hand the intention of the framers concerning the interpretation of the Convention. At a pragmatic level, the large number of people involved in the conferences and meetings culminating in the Convention makes it very *difficult* always to ascertain the objective of a particular provision, word or phrase; this has not, however, prevented courts and commentators from undertaking such an exercise: with varying degrees of apparent success. Thus, the task is not necessarily overwhelming. However, it is undesirable even to set upon this path. We now discuss two reasons why such intentions should be ignored.

## THE IRRELEVANCE OF THE FRAMERS' INTENTION: THE SOURCE OF THE AUTHORITY OF THE CONVENTION AND MAINTAINING THE RELEVANCE OF THE CONVENTION

### *The Authority of the Convention*

The interpretation of any legal document should be consistent with its source of authority.[56] It seems almost axiomatic that an interpretive approach inconsistent with the authority of the law may put at risk its legitimacy. The authority of the Convention has nothing to do with the framers. Absent ratification by state parties, the Convention

[55] For example, if a computer could be developed to measure desire content and intensity, group intention could be determined by the same program that ascertains individual intention.

[56] For example, see M S Moore, "Natural Rights, Judicial Review, and Constitutional Interpretation", paper presented at the conference "Legal Interpretation, Judicial Power and Democracy", Windsor Hotel Melbourne, 12–14 June 2000.

is irrelevant and its importance would be no greater than the piece of paper upon which it is written. The Convention only springs to life by virtue of its ratification and implementation into domestic law by nation states.

It follows that each party, as a consequence of its sovereignty, reserves the authority to place its "own slant" on the Convention. This can be done in two main ways. Either the legislature can directly alter the terms of the Convention so far as it applies at the domestic level, for example by severing, clarifying or adding to the international document. As we noted above, the Australian parliament has recently clarified many aspects of the Convention by amendments to the Migration Act. More subtly, a convention can take on a uniquely domestic flavour by virtue of the judicial interpretations that are given to it. This best example of this concerns the interpretation of the phrase "particular social group", where Australian courts have adopted an approach which is at odds with that in other jurisdictions.[57]

There are many reasons why a state may wish to alter the meaning or interpretation of an international law document. One of the strongest reasons is that the social and political conditions which gave rise to the document are no longer relevant. As is discussed in the following section, this is the situation with the Convention. There is, of course, a possible danger associated with this. If the instrument document is altered so drastically that its application no longer can be said to be in conformity with its terms, this can amount to a breach of international law and, in the case of the Convention, Australia might become subject to a dispute before the International Court of Justice (conditional on Australia consenting to the ICJ having jurisdiction to determine the dispute), a designated UN tribunal or committee; or at the very least could incur the condemnation of other parties to the Convention.[58] However, it is submitted that there is no prospect of this in relation to the changes that are proposed, which as discussed below will demonstrably make for a more humane application of the Convention.

*Preserving the Relevance of the Convention*

From the discussion above, it follows that it is not the case that regard *must* be had to the framers' intention. In the following section we make the stronger claim that, as a matter of policy, the intention of the framers should be ignored if the Convention is to remain even partially relevant in today's world.

As we indicated in the introduction, the Convention was purpose-built to meet the challenges of a different era. The problems it was designed to address are no longer relevant; or at least they are certainly not the main issues of our time so far as forced human migration is concerned. The Cold War is over and there are no longer (if there ever were) points to be scored by applying the Convention criteria on a case by case basis so as to provide a sanctuary for those wishing to defect to the "free world".[59] It has been suggested that the Convention is so out of touch that it is unlikely that governments would sign up to the same Convention in today's environment.[60]

So far as contemporary humanitarian law is concerned, it has been argued that the definition of a refugee is flawed in several respects. The main shortcoming concerns the Convention grounds. As they currently stand, the Convention grounds (race, religion, nationality, membership of a particular social group or political opinion) are unduly

[57] See generally R Germov and F Motta, *Refugee Law in Australia* (2003, OUP). See also P Dimopoulous, "Membership of a Particular Social Group: An Appropriate Basis for Eligibility for Refugee Status?" (2002) 7 *Deakin Law Review* 387.

[58] See, for example, Blay, Piotrowicz and Tsamenyi, above n 6, Ch 1.

[59] See Millbank, above n 5.

[60] *Ibid.*

narrow and ultimately arbitrary. The only manner in which refugee law can be made non-arbitrary is to remove the Convention grounds as the cornerstone of the definition and, instead, base the definition on the concept of deprivation and need, not the *reason* for the need. As has been previously noted, "the great issues of our day transcend party, region, religion and sex".[61] A person who will die from starvation due to a drought deserves refugee status more than one who risks being imprisoned for venting his or her political opinion. To offer protection to the latter but not the former is not only arbitrary; but also represents a gross distortion of the hierarchy of human needs and interests; one that is, in fact, not only flawed from today's perspective, but was equally so at the time of drafting.

The logical conclusion of this line of argument is that the Convention should be abolished altogether. Given its limits, there is some merit in this view, yet ultimately this would be undesirable. The Convention is important because it is a universal, humanitarian, international instrument that offers some guarantee that the fundamental rights of desperate people will be safeguarded. By and large most nations observe their obligations under the Convention. "The Refugee Convention stands out as a measure that offers substance and 'teeth' to the concept of internationally recognised human rights."[62] As is noted by James Hathaway, the fact that a state party which has jurisdiction over a refugee automatically owes that person core rights (especially protection against non-refoulement) is the strength of refugee law: "it ensures that few refugees fall through the cracks of the protection regime".[63] Given this and lack of momentum at the international level for a serious commitment to other humanitarian laws, the most productive approach in relation to the Convention is to make it as relevant as possible to today's era as opposed to abolishing it.

Ideally, this would result in a wholesale change to the definition of a refugee, where the criterion for help is the relief of pain and suffering, as opposed to membership of a certain race or group or adherence to a certain ideological belief. This would require a formal change to the definition, along the following lines:

> A refugee is a person who owing to: (i) the fact that his or her life is in peril as a result of lack of food, water or shelter; or (ii) a well-founded fear of having his or her physical integrity or liberty violated; is outside the country of his or her nationality and is unable to avail himself or herself of the relevant resources or protection of that country; or who, not having a nationality and being outside the country of his or her former habitual residence, is unable or unwilling to return to it.[64]

Given the slow pace at which consensus is reached at the international level and the lack of appetite expressed by most states for absorbing needy foreigners[65] a change along such lines is not likely to occur in the foreseeable future.

This should not, however, prevent one pursuing a "Plan B". In this regard, the next most desirable state of affairs is to "interpret" or apply the Convention in a manner

[61] J Jackson, "Measuring Human Rights and Development by one Yardstick" (1985) 15 *California Western International Law Journal* 453, 456–60.

[62] M Crock, *Immigration and Refugee Law in Australia* (1998, Federation Press) 163.

[63] J C Hathaway, "Refugee Law is Not Immigration Law", *World Refugee Survey* 2002, (OUP) 43.

[64] This definition is further explored in M Bagaric and P Dimopoulos, "Refugee Definition, Time for a Fundamental Rethink – need as the criteria for assistance" (2003) 9 *Canterbury Law Review* 268.

[65] Amnesty International has recently described refugee protection as the "black spot" in the European Union's rights ambitions (Amnesty International EU Office, Press Release, *Asylum Seekers in Europe: The Real Story Amnesty International Launches Europe-Wide Campaign*, 25 September 2001: http://www.amnesty.org) and is deeply concerned that the focus of the EU's asylum policy is overwhelmingly "on how to keep [refugees] out, rather than how to protect effectively people fleeing from war, civil upheaval and grave human rights abuses" (*ibid*). In a similar vein, Niraj Nathwani states that "refugee law is in crisis precisely because altruism has reached its limits. . . . We need to face the fact of donor fatigue": N Nathwani, "The Purpose of Asylum" (2000) 12 *International Journal of Refugee Law* 354, 356.

which will enable states to assist as many needy people as is possible. Ascribing weight to the framers' intentions – which were formed against the backdrop of a desire to assist those fleeing parts of Europe in the Cold War – will continue to frustrate this objective.

It is critical that the Convention is interpreted in a manner which provides the maximum possible assistance to those in need. Adherence to the intention of the framers will frustrate this goal and potentially lead to the document becoming obsolete (if it is not already). As the definition currently stands, second order human interests (such as the right to project one's political opinion) are prioritised over basic human needs. In our view, the definition should, as far as possible, be interpreted in a manner that enables the world's collective compassion to be targeted more directly at those who are suffering the greatest degree of deprivation.

This could be achieved with two main interpretive shifts. The first would see the grounds interpreted as widely as possible. The amorphous ground of "particular social group", in particular, provides scope for this to occur. Moreover, we note that one of the fashionable reform proposals in refugee literature at present is that women should be recognised as a particular social group.[66] We agree, and in fact would go further: men should be too. The effect of this is that being a member of the human species is the sole criterion that marks one out as being worthy of compassion. The *reason* for one's pain and destitution is not cardinal: pain and destitution suffice. It is, of course, unlikely that any court would be bold enough to interpret "particular social group" in such an open-ended manner. Nevertheless, it remains true that the wider the definition the better.

The other main interpretive shift that is required is to the notion of persecution. Historically, the notion of persecution has not been defined with any level of exactness.[67] As is noted by the UNHCR, "there is no universally accepted definition of 'persecution', and various attempts to formulate such a definition have met with little success".[68] Hathaway states that "persecution may be defined as the sustained or systematic violation of basic human rights demonstrative of a failure of state protection".[69] As noted above, in Australia, a key element of persecution is serious harm and the expression "serious harm" is defined (in an non-exhaustive manner) to mean a threat to life or liberty, significant physical harassment or ill-treatment; or significant economic hardship or denial of access to basic services or denial of capacity to earn a livelihood, where such hardship or denial threatens the applicant's capacity to subsist. This definition was in response to court judgments which significantly lowered the threshold concerning the types of harm that would qualify for refugee status.[70] For example, it was held that discrimination in the form of restrictions to employment and educational opportunities could constitute serious harm.[71] In order to

[66] See, for example, N Kelley, "The Convention Refugee Definition and Gender-Based Persecution: A Decade's Progress" (2002) 13 *International Journal of Refugee Law* 559; M Falcon, "Gender-Based Persecution" (2002) 21 *Refugee Survey Quarterly* 133.

[67] As is noted in the UNHCR *Handbook on Procedures and Criteria for Determining Refugee Status*, 14, "there is no universally accepted definition of 'persecution', and various attempts to formulate such a definition have met with little success".

[68] UNHCR, above n 67, 14.

[69] Hathaway, above n 12, 104–5.

[70] S Haddad, "Qualifying the Convention Definition of Refugee" in (2002) *Immigration Review* 101.

[71] See *Chan v MIEA* (1989) 169 CLR 379, 431. See further, *Syan v Refugee Review Tribunal* (1995) 61 FCR 284 (extortion); *Applicant A v Minister for Immigration and Ethnic Affairs* (1997) 190 CLR 225; *Thalary v Minister for Immigration and Ethnic Affairs* (1997) 73 FCR 437; 50 ALD 349; *Gunaseelan v Minister for Immigration and Multicultural Affairs* (1997) 49 ALD 594 (discrimination in employment and education); *Minister for Immigration and Multicultural Affairs v Gutierrez* (1999) 92 FCR 296; 59 ALD 89 (exposure to civil litigation).

constitute serious harm it was unnecessary for the applicant to be denied the opportunity of *any* employment; merely being denied the opportunity to work in a chosen field was sufficient.[72]

Persecution, unlike the Convention grounds, should be interpreted narrowly. In this regard, something of a paradox emerges. Ostensibly, an expansive interpretation of persecution may appear to assist the refugee cause by expanding the type of harm that qualifies for assistance. Viewed literally this is no doubt the case: the people or persons who come within the expanded definition are granted access to the relevant nation state. However, at this point good intention and good consequences part company as a result of the finite level of assistance that the international community is willing to provide to needy foreigners. Given the scarcity value and preciousness of refugee places, the kind thing to do is not to expand the range of human interests that are recognised under refugee law, but rather the opposite. The interests should be narrowed to ensure, as far as possible, that the refugee places are occupied by those in greatest need. A related point is that, while an expansive definition of serious harm appears to be consistent with the humanitarian underpinnings of the Convention, it is ultimately misguided because it verges on merging refugee law and immigration law. Refugee law is not about equalising the international playing field so far as the capacity for people to flourish is concerned, it is about assisting those greatest in need.

## CONCLUSION

At the international law level, the Refugee Convention is the most widely adhered to humanitarian document. Thus, it is critical that the Convention is interpreted in a manner that provides the maximum possible assistance to those in need. As we discussed above, adherence to the intention of the framers will frustrate this goal and potentially lead to the document becoming obsolete. This is because as the definition currently stands, second order human interests (such as the right to project one's political opinion) are prioritised over basic human needs. A Convention which does provide maximum assistance to those most in need, by giving priority to basic needs, could be achieved through two main interpretive shifts. The first would see the grounds for refugee status interpreted as widely as possible, in order to encompass individuals whose basic needs are not being fulfilled. The other main interpretive shift that is required is to the notion of persecution. In order, to achieve this, law-makers need finally to put to rest the intentions of the framers of the Convention.

[72] See, for example, *Ahmadi v MIMA* [2001] FCA 1070 (Wilcox J, 8 August 2001).

# ACCESS TO JUSTICE: EMPLOYMENT TRIBUNAL CONTINGENCY FEES, WHAT CHANCE OF JUSTICE?

*JANE JOHNSON* and GERALDINE HAMMERSLEY***

*Background and Introduction*

Industrial Tribunals, re-named Employment Tribunals by the Employment Rights (Dispute Resolution) Act in 1998, were created in 1964 under the Industrial Training Act, section 12 to provide a forum for resolving disputes between employers and employees; namely, at that time, disputes concerning the payment of the Industrial Training Levy. Further, under the Industrial Relations Act 1971, their jurisdiction broadened to include unfair dismissal claims and since then they have been empowered to hear claims on all aspects of employment legislation. Moreover, the continual introduction of new legislation regulating the employment relationship has given rise to an inevitable increase in the number of employment disputes and the number of applications being made to employment tribunals: a threefold increase since 1990, averaging over 90,000 applications per year and forecasted to reach 100,000 in 2004.[1] Not all these cases proceed to a full hearing at an employment tribunal but may be settled through conciliation or mediation.[2] For the future it is hoped that the introduction of a new statutory dismissal, disciplinary and grievance procedure, which came into force in October 2004, will reduce the amount of cases proceeding to tribunal, as the responsibility for dispute resolution in these areas is placed firmly in the hands of the employer.[3] Further, unlike actions in the civil courts, there is no general rule in employment cases that the successful party has the right to recover their legal costs. Employment tribunals have discretion to award costs in limited situations, for example, where a party or a party's representative has acted vexatiously, abusively, disruptively or otherwise unreasonably.[4] However, the general presumption still remains that costs will not be awarded, and thus those with uncertain claims have nothing to fear if they lose their case. Conversely, where a claim is successful, the no cost rule does prevent the applicant's lawyer being paid by the losing respondent. Moreover the increase and acceptability of contingency fee funding arrangements for legal representatives has meant that applications taken on this basis have nothing to lose. Under the contingency fee arrangement (not available in general litigation but not excluded in the employment tribunal), the client enjoys several advantages. For example, there will be no anxieties about having to pay large legal fees, and no requirement to pay before a case is heard. Also, the lawyer is deemed to be taking the risks and in most cases should have carefully assessed the chances of winning. The lawyer simply takes a percentage of the award made by the employment tribunal as agreed between the parties. In contrast, charging an hourly fee for legal advice does provide the most lucrative alternative for practitioners: lawyers rarely lose out on such arrangements. However, as such rates can be very high, averaging £125–£150 an hour, it would be very easy for claimants quickly to

*LLM, Senior Lecturer in Law, Coventry Business School, Coventry University.

**MA, Principal Lecturer in Human Resource Management, Coventry Business School, Coventry University.

[1] *Employment Tribunal Service Annual Report 2002–3* (The Stationery Office, London).

[2] There is a statutory duty on ACAS conciliation officers to endeavour to promote a settlement where possible, s 18 Employment Tribunals Act 1996.

[3] Employment Act 2002 Schedule 2, Part 1 and 2.

[4] Rule 14, Employment Tribunals (Constitution and Rules of Procedure) Regulations 2001 (SI 2001/117). See *Beynon and Ors v Scadden and Ors* [1999] IRLR 700 and *McPherson v BNP Paribas* [2004] EWCA Civ 569, [2004] 3 All ER 266.

run up heavy costs, hence the popularity of no win, no fee arrangements and other alternative cheaper ways of obtaining legal advice, through, for example, Citizens Advice Bureaux and trade unions.

Under conditional fees, a client who wins his or her case has to pay the lawyer's professional fee, and in addition a success fee, which is calculated as a percentage of the professional fees. If the client loses, he or she may have to pay the winning side's costs including the success fee charged by the successful party's lawyer and any insurance premiums paid out by the winner. However, there would be no professional fee charged by the client's own lawyer. Generally, these are most common in personal injury cases, as the winning party will usually be able to recover costs and for this reason they are rarely used in employment cases. Conditional fee arrangements were legitimised under the Courts and Legal Services Act 1990, section 58, as part of the Thatcher government's radical reform of the legal profession and the shift to free competition and choice. Both contingency fees (where available) and conditional fee arrangements may be described by the umbrella term "no win, no fee arrangement". In the context of the study described in this article, however, the term "no win, no fee" refers exclusively to contingency fees unless it is otherwise clear from the context.

The current funding arrangements for applicants seeking legal advice in employment matters have resulted from legislative changes; in particular, the Access to Justice Act 1999, which replaced the legal aid system in England and Wales with two new schemes: "Legal Help" and "Help at Court". These schemes are administered by the new Legal Services Commission, which replaced the Legal Aid Board. The Legal Services Commission has no power to provide financial help available for representation before employment tribunals. However, some solicitors and other bodies may give limited free advice to those with low disposable incomes as *pro-bono* work, or in special cases, "Legal Help" (although not for actual representation) through the Community Legal Service, the generic title used by the Legal Services Commission in relation to civil cases. The Community Legal Service (CLS) has developed a network of "Legal Help", through Citizen's Advice Bureaux, Law Centres, and law firms. Information on this network has been publicised through public library information centres and a website.[5] Additionally, the CLS, having recognised the importance of quality legal help, lists selected lawyers and advice centres with a Quality Mark, as a guide to standards for the general public. Employees who are members of a trade union may also have access to legal help and advice (some unions offer a 24-hour legal helpline) and unions can advise on the merits of particular cases. Additionally, the Equal Opportunities Commission, the Commission for Racial Equality, and the Disability Rights Commission may assist in the funding of a case where appropriate. Failing this the Free Representation Unit, which is a charity, provides free representation in the Croydon, London North and Staffordshire East Tribunals. The Bar Pro Bono Unit also offers free employment law advice and is assisted by the Employment Lawyers' Association and several solicitors.

Overall, there is little financial help available for those involved in employment disputes, and little guidance on where to find help and advice for funding and choice of legal representation.[6] Research undertaken to date demonstrates high levels of

[5] www.justask.org.uk.

[6] Now a recommendation of the ETS Taskforce 2003 (paragraph 5.26), Department of Trade and Industry DTI/6174/2K/07/02/NP. URN02/1091.

dissatisfaction with the current system; experiences of applicants suggest that the advice they received was poor, especially in cases where they received advice from non-qualified legal representatives.[7]

This article focuses on the fee and funding arrangements in employment law cases, the quality of the legal service provided, and whether there is a correlation between contingency fee and hourly paid arrangements in terms of quality and access to justice for those concerned. To meet these objectives, empirical research has been carried out in the form of a survey of both legal representatives and applicants. These results have been analysed and provide evidence from which conclusions and recommendations can be drawn.

*Method and Sample*

In order to discover empirical evidence of the realities of the no win, no fee phenomenon in employment law cases, questionnaires were distributed to a random selection of 800 law firms in the UK during November 2003. The sample was taken from the on-line Yellow Pages Directory, and most of those targeted did offer employment law advice. In particular their views were sought regarding fee arrangements, access to justice and the impact of such arrangements on their firms.

A further questionnaire was distributed to 1,000 applicants whose cases had been listed on the Public Register for hearing in a tribunal. This aimed to investigate the experience and perceptions of those involved in employment disputes, for example, the fee arrangements, the quality of representation and overall satisfaction with the legal advice and outcome. Face to face semi-structured interviews also took place with a sample of the applicants who self–selected for interview within their responses. These were conducted in February and March 2004.[8]

Questionnaire results were analysed using SNAP, a Survey and Analysis Package produced by Mercator, a leading specialist in processing survey questionnaires. Interviews were tape-recorded and transcribed. Subsequently, they were analysed using both stakeholder and content analysis. Stakeholder analysis[9] was utilised when analysing the interview transcripts as it allowed the researchers to consider data that was subjective and personal, including interviewees' opinions, experiences and perceptions from an individual point of view, as well as interpersonal/intersubjective opinions and perceptions that were shared by more than one informant. Following the identification of these opinions and perceptions, content analysis was undertaken. As the data was from a variety of individual sources and somewhat fragmented in the initial stages, content analysis afforded a methodical way of highlighting its complexity. Constructs and themes were identified from the transcripts, which were then categorised and counted for frequency of mention. This enabled the development of a matrix with constructs on one axis and frequency of mention on the other. Whilst the use of these techniques was lengthy, we considered it the most suitable way of dealing with the qualitative data.

[7] *The Blackwell Report* para.45–47, The Report of the Lord Chancellor's Committee to investigate the activities of non-legally qualified claim assessors and employment advisers (18 April 2004, Department of Constitutional Affairs).

[8] In all, 20 interviews took place.

[9] For a fuller discussion of stakeholder analysis see Burgoyne, J, in Cassel, C and Symon, G; *Qualitative Methods in Organizational Research* (1994, Sage Publications).

## FINDINGS AND ANALYSIS

*Survey of Legal Representatives*

There were 85 respondents; 56 of which offered contingency fees, and although a return rate of 10% seems low, sufficient information has been collected, through a detailed questionnaire, to form some interesting analysis. It is also well established that it is difficult to get large response rates when lengthy questionnaires are sent to lawyers.[10]

The majority of the firms used solicitors to represent clients at employment tribunals (84%). However, as respondents could tick more than one answer, the results showed that a minority used non-legally qualified consultants and barristers. The majority agreed the same representation would apply to cases funded by fee arrangements other than those calculated under traditional hourly rates.

The size of the firms in this exercise varied with those having less than 20 individuals working at the firm counting for 19% of the sample and those having more than 20 but fewer than 100 counting for 53%. Twenty-eight *per cent* had more than 100 workers.

The sample consisted of a representative cross-section of those representing employers and employees. Almost half the respondents represented both factions. Interestingly, of those who represented employers, 70% did not offer contingency arrangements at all and the remaining 30% only offered them in very exceptional cases: less than 25% of their workload. This can be contrasted with those who acted solely for employees, where 80% (18 out of 22) offered no win, no fee and of these, 40% (8 out of 18) offered this arrangement in over 50% of employment cases. This response is substantiated by a further question which indicated that 96% of the law firms offered no win, no fee arrangements when representing applicants.

In total, 66% of respondents (56 out of 85) offered no win, no fee arrangements. Where respondents did not offer no win, no fee arrangements the questionnaire invited them to explain why such arrangements were not offered. These comments varied. Some felt that the uncertainty surrounding the outcome of such cases meant that they were not a viable option. For example, one respondent said, "As a policy decision we consider employment tribunal work too inherently risky to offer conditional fees [sic]". Another stated that it "was not appropriate for respondent work, too risky". Also several felt that contingency fees were "not cost effective", and one claimed, "We start to lose thousands of pounds in potential costs as does counsel."

Several stated that they acted for employers who had the money to pay hourly rate fees, or in fact would be covered through insurance: "In the main, the work is employer based. The employers who are represented normally have sufficient means to discharge their own costs." Additionally, one respondent felt the scheme unsatisfactory "because clients are not putting up their own money so do not have the same interest in their case . . . and there are enough clients who will pay normal fees". Another respondent stated that, from experience, the firm had lost a lot of money in potential costs, which, of course, cannot normally be reclaimed in an employment tribunal. Some were clearly against this form of fee payment on moral grounds. One stated "it is undesirable for the lawyer to have a personal interest in the outcome", and another said, "The scheme is dishonest and award usually does not justify the fee." However, one may question whether this espoused ethical stance simply masks hard-nosed commercial reality and a preference for an hourly rate of pay.

[10] S Yarrow, *The Price of Success Lawyers, Clients and Conditional Fees*, (Policy Studies Institute, 1998) Results were analysed from 121 firms of solicitors. The fieldwork was carried out in 1996–7.

Almost half the sample who had been offering no win, no fee arrangements had been doing so for more than a two-year period and, when asked whether this arrangement was offered in all employment disputes, the majority replied negatively (88%). The reason for this gap was because assessment and arrangements were decided on a case-by-case basis. Several stated that fee arrangements were decided upon depending on the prospects of success in a case and the potential award. One respondent suggested that the firm would only take on a case if "their prospects of success are greater than 50% and if they are the applicant in the case". Another said, "We would not take on a case on no win, no fee with less than a 60% chance of success." Another stated that they would only take on such a case where "chance of winning was at 80% or more". Others commented that cases would only be taken on this basis "in limited circumstances" if there were "good prospects of success" and a reasonable award at stake, this being in the region of £15,000. Also some respondents stated that the client had a choice of an hourly fee or no win, no fee arrangement, and "it depends on the applicant's financial circumstances". Some individuals have private funds, whilst others are insured, in some cases on household policies, and others had nothing. However, it was interesting to note that, in cases where the legal issues were complex, such as discrimination, there was a general trend within the responses that a case would not be taken on a no win, no fee basis as opposed to an hourly rate, because of the time it would take. Several respondents indicated this, for example: "Not offered in complex cases, such as discrimination, because the amount of work required is disproportionate to the costs recovered". This comment has to be viewed with some scepticism as, when respondents were further asked to comment on their last employment law case, it was clear that many did operate a no win, no fee arrangement in discrimination disputes. It is also clear that many applications are made to employment tribunals under this head.[11]

Further, no win, no fee arrangements were not usually offered to employers, and the majority of respondent firms did carry out some form of risk assessment exercise before organising the fee arrangements and, of course, agreeing to represent the individual. These comments are further substantiated by responses to a question asking whether the fee arrangements were offered dependent upon the value of the claim. 75% replied "Yes" and 21% replied "Sometimes".

Seventy-four *per cent* of those offering no win, no fee arrangements stated that these arrangements provided, or at least sometimes provided, better access to justice to victims in employment disputes.

When questioned about the perceived benefits of contingency fees the majority of the respondents felt that there were major cost advantages to applicants. For example, those who otherwise would not be able to afford to have legal representation were enabled to enjoy this benefit if they had viable cases. The system basically provides for a wider access to justice and offers an alternative mechanism for funding. One respondent stated that it provided "access to justice for a few, but encourages spurious claims as no risk to client". In contrast, another stated that it "allows a client to pursue a claim they could not ordinarily afford. It also weeds out potential claims that will probably not succeed". Several of the respondents made clear that all the advantages were for the client; since, the "client does not have to pay if case is lost"; "less risk for clients"; "lowers money worries" and it "enables clients who can't afford representation to bring their case with a considerable greater chance of success".

[11] ACAS *Annual Report* 2001–02 (TACAS). There were 19,021 applications to employment tribunals claiming discrimination on the grounds of sex, race, disability and equal pay.

Additionally, two respondents felt the scheme brought in "more clients" and "attracted clients who would not necessarily be prepared to instruct a solicitor on an hourly basis".

These comments may be viewed with some scepticism as law firms would surely not offer no win, no fee funding arrangements unless they provided a viable income (unless it was possible to treat such work as a loss leader).

One problem with the no win, no fee arrangement lies in the fact that it is not particularly beneficial for those who have claims that are good, as far as the merits are concerned, but of low value. It favours those who have high value claims because the lawyers only want to take on those cases which will give them a good chance of financial success. This substantiates the findings above, where it is clear that some applicants are prevented from pursuing low value cases as legal representatives clearly do not feel the rewards of conducting the case on a no win, no fee basis would be worthwhile.

In relation to the way in which the firms would be affected if no win, no fee arrangements were not offered, the respondents could tick more than one answer. Forty-six *per cent* of those offering no win, no fee arrangements indicated that their workload would diminish, and one in fact stated that the firm would cease business. Thirty-four *per cent* stated that the firm would not really be affected, and 50% agreed that if no win, no fee arrangements were not offered, then clients would lose their access to justice.

When questioned about the way in which the contingency fee was calculated, 60% said that it was usually between 25 – 50% of the client's award. None stated more than 50%. One respondent stated that around 33% (inclusive of VAT) of the award was usually charged to the client. Some also varied their fees if cases were settled prior to a hearing. For example, 25% if settled before the tribunal hearing and 40% if it went to a full hearing. Further, when respondents were asked whether their clients would have to pay additional disbursements, 40% of those offering no win, no fee arrangements indicated that clients may have to pay court fees, expert's fees and counsel's fees. In the "other" category, some respondents commented that expenses might include travel costs, medical reports or other evidence.

Sixty-two *per cent* of respondents felt that their work load had not increased as a result of no win, no fee funded claims, and 54% felt that the existence of a no win, no fee arrangement never placed them under pressure to accept early settlement of claims, although 46% felt it sometimes did. Likewise, 52% also felt that the existence of a no win, no fee arrangement never made them feel under pressure to take a case to full hearing, although 46% replied that it sometimes did.

Law firms were asked whether they would prefer their representatives to charge a set fee for their work, rather than offer the no win, no fee option. Fifty *per cent* of the sample said "yes". They were then asked to explain why they would prefer to charge a set fee, such as an hourly rate. The explanations followed a theme based on commercial reality, similar to that expressed above in relation to the lawyer's need to take on viable cases to ensure financial gain. Several felt that the hourly rate was preferable as it was a "guaranteed income" for their law firm; "certainty that you will get something back from the case" rather than "the lottery of the tribunal". The no win, no fee system in employment disputes tended to work out cheaper for the applicant and one respondent stated that his firm "tended to lose money" under contingency fee arrangements. From the client's point of view, some respondents felt that the no win, no fee system did not offer proper access to justice because "compensation limits are low and often just represent the loss sustained. Any costs

later taken out of that means the aggrieved party is inadequately compensated". Additionally, legal representatives often do not take on these cases "unless they do the work in bulk" and this does not always provide the best legal service to victims in employment disputes: "On the whole, my opinion is that contingency fee arrangements are not really an option for those seeking justice." Further, it was mentioned that where a fixed rate is charged, then there is never a concern over whether the case might be settled earlier or whether it should go to full trial.

In order to provide further substance to this research, questions were asked under conditions of confidentiality about the legal representative's last employment law claim under the no win, no fee scheme. Seventy-one *per cent* of the respondents stated that their last case was an unfair dismissal dispute.[12] The rest were discrimination or constructive dismissal cases. Forty-six *per cent* of those cases went to a full hearing; 54% did not. Eighty-eight *per cent* of respondents claimed that they felt their client was satisfied with the outcome of their case whilst 64% of the sample felt that they had achieved an excellent result for their client.

It is impossible to draw any major conclusions from these responses as there are difficulties in measuring levels of satisfaction, and one would expect lawyers to protect and promote their professional prowess.

When questioned over what fees were charged, 48% stated that they charged their client a success fee, and "other" charges included disbursements, such as travel and medical reports, or an agreed percentage of the award in contingency fee cases. The amounts varied between 30% and 42% of the compensation awarded. Further charges were also made in some cases for counsels' and for experts' fees.

The length of time spent on the respondent's last case in preparation and attending the hearing varied from less than half a day to more than five days. Overall 70% of respondents spent over two days working on the case.

Finally, respondents who offer no win, no fee arrangements (56 out of the original 85 respondents: 66%) were invited to make additional comments regarding funding arrangements. The following points were made and are worthy of note, in spite of the limited sample. Several comments suggested that contingency fee arrangements were not effective from both the lawyer's and client's perspective. For example, from the client's point of view, a case would not generally be taken on by a lawyer on this basis unless there was a good chance of a high award: generally only claims likely to achieve £12,000–£15,000. Consequently, employees who are victims claiming smaller amounts may have to settle their legal fees on an hourly rate basis, which may easily result in little financial gain. Another respondent felt that such a scheme seemed "an unfair burden on the client", since, where a high award was achieved, the lawyer might gain more than substantial fees. Additionally, some clients might pay more for the legal service than they would have done if they were paying an hourly fixed fee. Conversely, one respondent commented that in some cases lawyers will offer no win, no fee arrangements as the bulk of their work and thus operate on the basis of "win some, lose some". This may put more pressure on the lawyer to take on more cases and, as these findings suggest, settle as quickly as possible. This may or may not be advantageous to the client as there could also be non-financial advantages to the applicants to settle the matter quickly.

Another respondent stated that such arrangements are "fraught with danger: they can encourage early settlement which might sell the applicant short; if the respondent legal representative finds out, it may deliberately run up cost unnecessarily". This

[12] An expected response, as unfair dismissal claims form the bulk of employment tribunal work. 45,373 applications were made in 2002/3 to both ET and EAT (*Employment Tribunals Service Annual Report 2003*).

statement is substantiated by another respondent who commented that such arrangements "are dangerous to offer in tribunal cases unless solicitor has an eye for winning the case at a very early stage". Further, one respondent claimed that it was difficult to assess whether a case should be taken on a no win, no fee basis, and on no account would a trainee or junior fee-earner be allowed to take such a decision. Additionally, the fact that the tribunal system tended to vary considerably in terms of speed and efficiency, together with the problems of adjournments meant that some no win, no fee arrangements could be highly uneconomic. Further criticism of the tribunal system was also registered by one respondent who had had experience of the tribunal allowing witness evidence beyond the agreed timetable and in, one other case, two new and undisclosed witnesses were allowed to give evidence three quarters of the way through the proceedings.

In contrast, to the above, another respondent stated that he or she had

> several criteria to apply to such cases ... I never take them on simply in the hope of settling and do not take them on as "bulk" work. We pride ourselves on excellent service for every client and have a high success rate on conditional fee cases.

*Survey of Applicants*

A questionnaire was distributed to 1000 applicants whose names had been selected from the Public Register and were awaiting an employment tribunal hearing. There were 99 responses, just over 60% of these were male, and the majority were aged between 26 and 55 years old. Their occupations were diverse, ranging from unskilled, skilled and semi-professional to professional. Forty-eight *per cent* were from professional or managerial backgrounds, and 29% were from the manual sector. Just over half of the applicants were in disputes over unfair dismissal, and almost 25% had a claim under discrimination laws. When the respondents were asked who was representing them, 33% stated that they were representing themselves. Twenty-eight *per cent* had solicitors representing them, half of whom were employed on a no win, no fee basis, the others on a traditional hourly rate. Twenty-one *per cent* were represented by their trade union.

Only 27% of the respondents had had their case heard at the time the questionnaire was distributed: the remainder were still awaiting a hearing or were continuing negotiations. However, by the time the face-to-face interviews were conducted, all interviewees' cases had been heard or had reached a settlement.

Generally, questionnaire respondents who had been represented were satisfied with the way their case had been handled but some made suggestions about how they would have liked the case to have been handled differently. These included comments such as being kept more informed of the progress of the case, and also greater efforts being made to speed up what tended to be a lengthy process. Several examples were cited where individuals felt they had been pressurised to settle their claims earlier than they would have wished. A minority of respondents also suggested that their representative did not have the specialised knowledge to deal with the case effectively.

The results of the face-to-face semi-structured interviews showed that all of those who participated in the interviews were dissatisfied with the tribunal system. Whilst this might explain their willingness to participate in the research, they did nevertheless provide some interesting views of the no win, no fee system. In particular, content analysis highlighted the following themes emerging as key issues for applicants: feeling pressurised into accepting settlements, not being kept informed by their representatives, not understanding what was going on with their case and ultimately that pursuing a claim was stressful and they would not wish to go through it again.

The majority of applicants represented on a no win, no fee basis felt pressurised into accepting settlements. This was particularly prevalent in cases when they had managed to obtain alternative employment fairly quickly:

> I thought I had a very strong case and I thought perhaps it could have been pushed a bit harder with regards to the actual award . . . but he said this is what we class as a peanut case. I knew exactly what he meant, because I had got another job so quickly.

Interviewees felt that they were not kept informed of the progress of their case. In one instance a respondent suggested her enquiries regarding progress of her case were unwelcome intrusions for her solicitor: "every time I contacted him I was made to feel like I was mithering him". Additional problems were encountered by respondents when their legal representatives were changed at short notice and without prior discussion:

> The only issue I had with my solicitor was I actually sourced a lawyer first for the initial consultation and that is why I went with that firm and then they changed me to his assistant who I didn't think was very good and didn't instil much confidence in myself, that whizzed through and then she left and I had another assistant. So there was a bit of changing around between the people themselves.

All participants commented on the stressful nature of pursuing a claim at tribunal and that ultimately the funding arrangements were an additional cause of stress. Worries about legal costs were cited as a prime source of stress, alongside fears of getting into debt should they lose and have costs awarded against them.

> I was thinking because I want compensation, because they treated me unfairly, because obviously I am not going to be able to afford £110 an hour representation fee, so I thought well I can't really take this any further.

## CONCLUSION

In conclusion, no win, no fee arrangements have arisen as a result of the reduction in legal aid and the economic reality of the UK system of justice. The UK was in fact the first of its European associates to introduce conditional fees, followed by Belgium and the Netherlands.[13] In terms of access to justice, this research can only provide an analysis of the process and experiences of the lawyers and litigants. It fails to measure how many applicants are deterred from bringing claims in the first place. Certainly, some of the disadvantages listed may easily amount to reasons why an employment dispute might not be pursued in the first place. For example, Genn found that when she focused upon employment disputes, 16% of individuals "had done nothing" about their problem. One of the main reasons offered was because the individual thought that nothing could be done, or that the employer was right. In two *per cent* of the cases, the individual felt it would cost too much, and in four *per cent* of cases, it would take too much time.[14] The latter two problems have clearly been identified as serious issues for applicants within this research project.

Although analysis of a sample of 56 legal representatives offering contingency fees and views of 99 individual applicants is somewhat limited, some important final conclusions can be drawn. Firstly, no win, no fee arrangements are of greater benefit to employee applicants rather than employers. This is due to the fact that employers are usually either insured for legal work or have the necessary funds to meet legal costs.

[13] W Emons, *Conditional versus Contingent Fees* (2004, University of Bern and CEPR).

[14] H Genn, *Paths to Justice*, (1999, Hart Publishing) at p 43.

It also emphasises the fact that the no win, no fee system is in existence to provide a legal service for those who otherwise may not be able to afford it, thereby broadening access to justice. Further, as recognised in the US experience, contingency fees may provide an incentive and encouragement for lawyers to take a case on, which, under the traditional hourly rate of pay scheme, they would have rejected. The same, of course, is true of the applicants themselves.[15] Conversely, it is clear that the no win, no fee arrangement is not always popular from the lawyer's point of view. The main reason for this is that the decision to take on a case on this basis is sometimes risky and could turn out not to be cost effective. This factor is substantiated by literature available regarding the US contingency fee system, which indicates that in a large proportion of contingency fee cases, lawyers actually make substantially less on an hourly basis than they would if charging a fixed hourly rate.[16] Some respondent lawyers perceived claimants as having all the advantages under such arrangements with the advisers taking all the risks, despite increasing their number of clients. It is submitted that the acceptance and wider use of contingency fee arrangements in employment tribunals have broadened access to justice although it is clear that the quality of that justice is sometimes threatened. For example, lawyers have to decide upon their chance of success in a case before accepting it. Varying responses of somewhere between 50% to 80% suggest that a claimant might be turned down by one legal representative but accepted by another. The chance of justice on this basis may depend on how willing the client is to test the market, yet clients do not generally shop around[17] and although they have the right to challenge a solicitor's bill, rarely do.[18] Thus, it seems that applicants will only be represented not only if the chance of success is good, but also if the likely award is more than £12,000. The smaller claims, although often of merit, fall by the wayside unless the applicant can pursue the claim through the network of "Legal Help" developed through the Community Legal Service. This problem is heightened by the fact that awards in employment tribunal cases are generally quite low, averaging £6,776 in unfair dismissal cases in 2003.[19] One may question the professional ethics of this situation where the offer of advice is contingent on the monetary value of the potential result, and varies between legal representatives as, where a fee depends upon speculation, it has been recognised as neither benefiting the administration of justice nor in the public interest.[20] However, this view is now out of date: there are many advantages for both lawyer and litigant and those who are more sympathetic recognise the new era of free enterprise in the legal profession and the commercial reality of running a law firm.[21]

Further, as contingency fees are not specifically regulated, firms are free to charge different percentages of the award made: generally, between 25–50% of the award. The claimant's task of finding the best possible legal representative who will charge the least amount for his or her services is more than difficult and can be likened to that involved in negotiating any commercial bargain. An individual may have his or her chance of justice, yet may lose much of his or her award through legal representatives who do

[15] H Kritzer, "Lawyer fees and lawyer behaviour in litigation: what does the empirical literature really say?" (2002) 80 *Texas Law Review* 1943.

[16] H Kritzer, "The Wages of Risk: the Return of Contingency Legal Practice" (1998) 47 *De Paul Law Review* 267, 291–293.

[17] S Yarrow and P Abrams, "Nothing to lose? Clients' Experience of using Conditional Fees" (1999, London, University of Westminster) Summary Report 5, (2000) Final Report, 25.

[18] M Zander, "Will the Revolution in the funding of civil litigation in England eventually lead to contingency fees?" (2003) Clifford Symposium De Paul University, Chicago April 18–19.

[19] *Employment Tribunal Service Annual Report 2003, op cit*, Appendix 1.

[20] D Tompkinson, "US Contingency Fees" (2002) March 152 (7025) *New Law Journal* 29.

[21] J Peysner, "What's wrong with Contingency Fees?" (2002) 10(1) *Nottingham Law Journal* 22.

not always follow the average 33% of damages success fee. One interviewee within this research project brought a case of disability discrimination against his employer, but gained very little in damages as his representative took a 50% contingency fee. This was an exceptional case, where the particular law firm had been recommended by the Citizens' Advice Bureau. There is no substantial evidence to show this is common practice.

Additionally, the results indicate that both legal representatives and applicants admitted to feeling pressured to settle a case quickly. It is clear that the legal representatives gain more in no win, no fee cases if they settle early, as time is money. Obviously, where hourly paid rates apply, the legal representative is content for his or her advice and negotiation to be ongoing. The issue of under-settlement is most interesting and this project has been unable to locate other empirical evidence by way of comparison. Peysner highlights the problems and difficulties of measuring under-settlement as individuals have "different expectations" of outcome, and in some cases, a faster result, although under-settled, could be more advantageous to the client.[22] However, settlement does operate against judicial precedent, on which the UK legal system is founded, and for this reason requires further investigation and research.

This project has demonstrated that there are advantages and disadvantages of the "no win, no fee scheme" in employment cases. They are commonplace and may in the future provide a role model for contingency fees to operate in other areas of litigation as opposed to conditional fees. It may also be better for litigants if more regulation were to be introduced relating to the percentage of award charged as a fee. Although 33% is the norm, this is not always the case. Furthermore, it is clear that applicants need extensive support and guidance when bringing cases against their employers. This could be provided through better communications from their representatives and if this were combined with a greater awareness of the legal advice and the variety of fee arrangements available then it may contribute to a decrease in levels of stress experienced and improve their chance of justice.

There is a plethora of employment legislation protecting individuals in the workplace, but as access to justice is limited to "good" claims, many deserving claims fall. The tribunal system was established to provide quick and speedy access to justice. The complexities of both the legislation and the formality of the tribunal system make it very difficult for individuals to represent themselves even though this was the original intention. Consequently, access to legal representation is vital. Additionally, a taskforce is currently reviewing the employment tribunal system with the aim of creating a system which "is fit for purpose for the 21st century" and to ensure that it can deal with cases in a just, fair and proportionate manner.[23] One may argue that until more regulation is in place and appropriate funding arrangements can be made for legal representation in smaller claims, many victims of workplace disputes may continue to lose their chance of justice.

[22] *Ibid.*

[23] Employment Tribunals Systems Taskforce, www.dti.gov.uk/er/individual/taskforce.htm.

# THE PROTECTION OF DESIGNS: A NEW RÉGIME?

*GARY SCANLAN**

## INTRODUCTION

The ambivalent and rather fluid relationship between copyright as a means of protecting designs and the various design rights which protect designs from unauthorised infringement has, it is submitted, never been satisfactorily resolved in English intellectual property law. This article will seek to define and delineate this relationship in view of both recent statutory developments and case law.

The case of *Lambretta Clothing Co Ltd v Teddy Smith (UK) Ltd,*[1] in particular, has given fresh impetus to a consideration and review of the means by which the law of intellectual property protects industrial designs that are essentially functional in purpose. The case therefore presents an opportunity to re-examine, albeit in a critical fashion, the relationship between design right[2] and copyright in so far as they may subsist in industrial designs. This article will address these matters, in addition to considering, albeit briefly, the implementation of recent legislative enactments relating to both the registered and the unregistered design régimes,[3] the effect of which may be, *inter alia*, to redefine the role and function of the national design right created by Part III of the Copyright Designs and Patents Act 1988.[4]

Before these issues can be addressed consideration should be given to the decision in the *Lambretta* case at first instance and subsequently in the Court of Appeal.

## THE *LAMBRETTA* CASE: THE FACTS

The claimant company was, and is, a designer and producer of fashion clothes for young people. In June 2000 one of its directors designed a track top (the Lambretta track top). The design had been recorded in two design documents, which collectively recorded in drawings and written instructions the nature of the track top. The design included, *inter alia*, the arrangement of colourways in the fabric of the track top. In 2002 the first defendant imported and sold a track top, which had been designed and produced in 2001 in France by a company related to the first defendant. The second defendant, Next plc, independently designed a track top having an appearance similar to that of the claimant's and this track top was first sold by the second defendant in England in 2002. These track tops were both similar to that of the claimant's design.

*Solicitor, Reader in Law, City University. I would like to thank the referees for their comments upon this article.

[1] [2003] EWHC 1204 HC, [2004] EWCA Civ 886 CA, 2003 WL 21353286 (Ch D), [2003] RPC 41.

[2] *ie*, unregistered design right as prescribed under Part III of the 1988 Act.

[3] Particularly the Registered Designs Act 1949 as amended by the Registered Designs Regulations 2001 SI No 3949 which came into force on 9 December 2001. See also the Community Design Regulation Council Regulation (EC) No 6/2002 of December 2001.

[4] However, a full and critical analysis of both the unregistered and registered design régimes created by the recently enacted Community Design Regulation Council Regulation (EC) No 6/2002 of December 2001 on Community Designs requires a separate article. Furthermore, in *Lambretta* both at first instance and in the Court of Appeal, the issue of the nature of infringement of design right was extensively discussed. Space and the balance of the article precludes a consideration of this issue, which also requires a separate and major article in its own right.

The claimant alleged that each of the defendants' track tops infringed its design right in its own track top,[5] alternatively that there had been infringement of the artistic and literary copyright that had been created in their track top by virtue of the design documents.[6]

## THE NATURE OF THE INTELLECTUAL PROPERTY RIGHTS CREATED BY THE CLAIMANT'S DESIGN

One of the principal issues in the *Lambretta* case was the application of the Copyright Designs and Patents Act 1988, section 51(1), and the degree to which copyright subsisted in the design.[7] However, the central aspect of Etherton J's judgment concerned the means by which designs such as the Lambretta garment could be protected by the law of intellectual property, and principally by design right.

### *The Nature Of Design Right*

Since the claimant sought protection of its design, *inter alia*, through the medium of unregistered design right, Etherton J summarised the nature and purpose of design right as created by Part III of the Copyright Designs and Patents Act 1988 (the 1988 Act).[8] Prior to the passing of Part III of the 1988 Act, copyright had been a principal recourse on the part of the creators of functional industrial designs as a means of protecting their designs from unauthorised reproduction or copying where such designs were not capable of registration under the Registered Designs Act 1949. Litigation had established the nature and extent of this protection in cases such as *LB Plastics v Swish Products*.[9] Any restriction on this form of protection in respect of functional industrial designs was only recognised in the comparatively recent case of *British Leyland Motor Co v Armstrong Patents Co*,[10] a decision of the House of Lords, delivered when the enactment of the design right régime under Part III of the 1988 Act was on the horizon.[11] The enactment of Part III of the 1988 Act, which created the unregistered design right régime, was in part modelled on the provisions contained in the Council Directive on the Topography of Semi-Conductor Products, which was itself implemented virtually word for word by the Semi-Conductor Products (Protection of Topography) Regulations.[12] The provisions of Part III of the 1988 Act were intended to provide a means of protecting functional designs intended to be produced industrially when copyright protection was removed from the realm of such designs by the 1988 Act.[13]

[5] *ie*, unregistered design right.

[6] Which consisted of both drawings and written instructions as to the nature and form of the design.

[7] This issue, although an important aspect of the case, is not considered fully in this article, space and the balance of the article precluding a detailed discussion of it. See, however, Scanlan "The Future of Design Right: Putting S51 of the Copyright Designs & Patents Act 1988 in its place" to be published in [2005] SLR, December, where this matter is extensively considered.

[8] [2003] RPC 41, para 25 *et seq*. In doing so his Lordship cited extensively from Laddie, Prescott and Vitoria, *The Modern Law of Copyright* (Sweet & Maxwell 3rd ed, year?) at para 53·1-53·3.

[9] [1979] RPC 551, HL.

[10] [1986] RPC 279.

[11] A case which concerned the claim for copyright protection in respect of designs for spare parts in the automobile industry, which was vigorously rejected by their Lordships, and which to a considerable degree was endorsed by the enactment of the must fit, must match exceptions to protection for such aspects of a design under Part III of the 1988 Act.

[12] Council Directive 87/54/EEC. It should be noted that the pre-existing topography right became part of design right by virtue of the Design Right (Semi-Conductor Topographies) Regulations 1989 SI 1989 /1100, which came into effect at the same time as the bulk of the 1988 Act.

[13] For a consideration, albeit briefly, of this aspect of the 1988 Act by reference to section 51 of that Act, see below.

## PREVIOUS AUTHORITY

Etherton J, in supporting his analysis of the nature of design right, cited the case of *Farmers Build Ltd v Carier Bulk Materials Handling Ltd*[14] where Mummery LJ also examined the background to design right. His Lordship in this case had noted that the recognition of copyright in industrial designs that were functional in purpose and conception had resulted in the protection of these forms of industrial designs for the full period of protection afforded to works of artistic copyright. This was not regarded as satisfactory, either from an intellectual property law or competition law perspective. The Whitford Committee,[15] despite reviewing the law of copyright and design, did not reach a unanimous and unambiguous conclusion on the appropriate régime to protect industrial or rather functional industrial designs. Nevertheless, as Mummery LJ further noted, following the White Paper on Copyright and Design and the enactment of Part III of the 1988 Act, it had generally been considered that the prime purpose and function of design right was to reduce the extent of the protection from copying afforded to the designs of industrially-produced and functional articles, and to remove such protection completely in respect of designs for spare parts. The full nature and ambit of design right therefore needed to be expounded by the courts. This, in part, is what Mummery LJ sought to do in the case of *Farmers Build.*

*The Nature And Ambit Of Design Right Following Farmers Build*

Mummery LJ regarded the nature of design right not as an amalgam of registered design and copyright, but more in the nature of a right with characteristics similar to copyright. This was based on the principle that the nature of the right afforded protection to the creator or owner of the design by way of an exclusive right to take action against third parties who, without authorisation, copied the relevant design. He noted that the purpose of unregistered intellectual property régimes such as design right, and (where still applicable) copyright was not to protect the "novelty of the work against all competition, but to provide limited protection against unfair misappropriation of the time, skill and effort expended by the author of the design on the creation of the work."[16]

## DESIGN RIGHT AND COPYRIGHT COMPARED: INFRINGEMENT

Apart from the relatively limited period of duration of protection prescribed for design right, as compared to that recognised in the case of copyright protection, the nature of infringement of a design subject to design right, according to Mummery LJ, was also restricted, since infringement could only take place in respect of a design where it was reproduced exactly or the reproduction was substantially similar to the original design, but only in respect of the design when considered as a whole. This can be contrasted with infringement of works protected by copyright, where infringement can take place where the work or any *part* of the work is reproduced exactly or substantially reproduced.[17]

[14] [1999] RPC 461, CA.

[15] Cmnd 6732.

[16] [1999] RPC 461 at p 480–481.

[17] It should be noted here that the test of originality for determining the aspects of a design which may attract design right and the nature of infringement of that right, although substantial and important parts of the *Lambretta* case at first instance and in particular in the Court of Appeal judgments are not considered in this article. Space and the balance of the article precludes such discussions. These issues, however, being important, require their own article.

## REGISTERED DESIGN AND DESIGN RIGHT COMPARED

Mummery LJ also compared registered design with design right from the point of view of the purpose and rationale of these two design protection régimes. Any design protected under the Registered Designs Act 1949 did not enjoy protection with regard to the aspects of the shape and configuration of the design which was dictated solely by the function of the design.[18] In his Lordship's opinion these aspects of a design in so far as they are functional in purpose are to be protected by the régime prescribed under Part III of the 1988 Act.[19] With this comprehensive analysis Etherton J agreed.

Etherton J, after citing the *Farmers Build* case, turned to a central substantive issue in the *Lambretta* case, namely the aspects of a design which were capable of protection under the design right régime. Design right as a means of protecting functional designs was capable of subsisting in aspects of the shape and configuration of the design in so far as those aspects were functional in purpose: such was to be considered axiomatic. What however, was the nature and relationship between these aspects of a design, and how did they relate to the claimant's design of the track top?

## THE SHAPE AND CONFIGURATION QUESTION: THE JUDGMENT OF ETHERTON J

It had been the common perception of intellectual property lawyers before the *Lambretta* case that shape and configuration[20] were synonyms within the design right régime, and that these concepts related to aspects of a design that would be represented in the finished article in *three* dimensions, even though design right in this context could exist both with regard to the external or internal parts of the design.[21] This view was supported, according to Etherton J in *Lambretta*, by authority such as *Mackie Designs Inc v Behringer Specialised Studio Equipment (UK) Ltd.*[22]

[18] But see below where the amendments to the Registered Designs Act 1949 since the judgment of Mummery LJ, may affect this rather prosaic statement of the nature and ambit of the registered design régime, and the relationship between registered design right and design right.

[19] This view may no longer be supportable without some qualification following amendments to the 1949 Act. This issue is considered briefly below, but is addressed at length in Scanlan and Gale "Industrial Design and the Design Directive: Continuing and Future Problems in Design Rights?" [2005] JBL 91–112.

[20] Which, by virtue of the Copyright Designs and Patents Act 1988, s 213, are the aspects of a functional design which are capable of being protected under Part III of the 1988 Act. The relevant parts of the section provide that:

"(1) Design right is a property right which subsists in accordance with this Part in an original design.
(2) In this Part 'design' means the design of any aspect of the shape or configuration (whether internal or external) of the whole or part of an article.
(3) Design right does not subsist in –
(a) method or principle of construction,
(b) features of shape or configuration of an article which
(i) enables the article to be connected to, or placed in, around or against, another article so that either article may perform its function, or
(ii) are dependent upon the appearance of another article of which the article is intended by the designer to form an integral part, or
(c) surface decoration".

[21] This was because a design protected by Part III of the 1988 Act is regarded as a design which is functional in nature and therefore lacking any aesthetic element or requirement. In these circumstances the internal aspects of such a design could therefore be protected equally with the external elements of the design which were open to ocular inspection. This position should be contrasted with the position of registered designs protected under the provisions of the Registered Designs Act 1949, which are regarded as designs produced for their aesthetic appeal, and which must therefore be capable of ocular inspection, but see below for a possible qualification to this statement of principle.

[22] [1999] RPC 717. Aspects of this case are considered in outline below.

*Shape And Configuration – Distinct Concepts?*

Counsel for the claimant in *Lambretta*, however, argued that configuration was in essence distinct from the concept of shape within a design and could include two dimensional arrangements of the design that included the colourways in garments such as in the Lambretta track top. The opposite and more traditional submission was made by counsel for the first and second defendants. In the defendants' submission, irrespective of the relationship between configuration and shape, configuration could only relate to arrangements in a design that were intended, when the design was produced as a finished article, to exist in *three* dimensions. Accordingly, configuration as a concept could not include aspects of a design of a garment which, when reproduced as a finished article made to the design, existed in only two dimensions. On the basis of this submission the aspects of the Lambretta track top design which consisted of colour patterns or colourways did not constitute configuration.

## THE *MACKIE* CASE

Counsel for the claimant in support of his submission had nevertheless, cited the *Mackie* case. He maintained that the case, which concerned infringement of the artistic and literary copyright subsisting in a circuit diagram of a piece of electrical equipment called a mixer, supported the contention that configuration applied to the aspects of a design which existed in two dimensions. The plaintiff in this case had alleged that the second defendant, having obtained an example of the mixer and analysed its circuits, had produced a "net list" of the components from which it then produced layouts for circuit boards that were manufactured by the third defendant. In considering the question of infringement of copyright, Pumfrey J had to consider whether the plaintiff's circuit diagrams constituted a design document within the terms of the 1988 Act, section 51. This section, which seeks to determine, *inter alia*, the situations where copyright may still subsist in the aspects of a design set out in a design document, defines the concept of design in essentially similar terms to the concept as defined in the 1988 Act, section 213,[23] including the requirement that the protection afforded to a design relates to the configuration or shape of the article the subject of the design. Pumfrey J considered the nature of configuration within the terms of section 51 and thereby, by analogy, the nature of the concept within the context of design right generally. His Lordship did so by reference to a definition of the term in *Webster's Dictionary*. The definition stated that configuration was constituted by a:

> a) Relative arrangement of parts or elements as (1) shape (2) contour of land (3) functional arrangement;
> b) Something (as a figure, contour, pattern or apparatus) produced by such an arrangement.

Pumfrey J held that the plaintiff's circuit diagrams were design documents within the meaning of the 1988 Act, section 51, because they showed the relative arrangement of parts or elements of the device.

Nevertheless, Etherton J in the *Lambretta* case rejected the argument of counsel for the claimant, that the adoption by Pumfrey J of the *Webster's Dictionary* definition of

[23] Which has been noted above at fn 20.

configuration in *Mackie* supported the contention that the concept of configuration included or involved aspects of a design which, if reproduced in a finished article made to the design, could exist in only two dimensions. In the words of his Lordship:[24]

> In my judgment, neither the quotation from *Webster*, nor the reasoning and decision of Pumfrey J in *Mackie*, support the primary case of Mr Campbell[25] that the mere juxtaposition of colours in two dimensions is sufficient to fall within the term "configuration" in s 213(2). First, the exposition in *Webster* referring to shape, contour of land, and functional arrangement clearly point to something, in the context of an article, other than the juxtaposition of patches of colour within the confines of an article.

Furthermore, his Lordship noted [26] that the whole tenor of Pumfrey J's judgment in *Mackie* was concerned with, *inter alia*, three dimensional components of a circuit board, accordingly, if anything, the case supported the orthodox view that design right protects the aspects of a design that are intended to be produced as part of an article in three dimensions.

Finally, counsel for the claimant in *Lambretta* sought to argue that, since surface decoration is expressly excluded from the ambit of design right,[27] this supported his submission that configuration and shape must include aspects of a design which, when reproduced in a finished article made to the design, would be in two dimensions. Etherton J rejected this submission based in part on the citation of Jonathan Parker J in the case of *Mark Wilkinson Furniture Ltd v Woodcraft Designs (Radcliffe) Ltd*,[28] where his Lordship had stated that surface decoration involved decoration on the surface of an article, for example a painted finish or decorative features on the article itself, examples include beading and engraving. Accordingly, in the opinion of Etherton J the exclusion of surface decoration from design right is not inconsistent with any restriction of the concepts of configuration or shape to the three dimensional elements of a design.

*The Defence Submissions*

Counsel for the defendants in *Lambretta* sought to use the Registered Designs Act 1949, section 1, which, as then drafted, protected any design that was an aesthetic design and which consisted of elements of configuration or shape in addition to pattern and ornament that exhibited this feature or attribute.[29] Counsel, however, cited this provision in order to reinforce the argument that configuration and shape within design right referred to those aspects of a design which consisted of three dimensional elements, on the basis that the terms must be interpreted in the same way in both sets of design régime. Etherton J also referred to the dicta of Luxmoore J in the case of *Kesto Ltd v Kempat Ltd*[30] where the learned judge, in considering the design elements necessary for a design to be eligible for registration under the then registered design régime,[31] had noted that:

[24] [2003] RPC 41 at para 39.

[25] Counsel for the claimant.

[26] [2003] RPC. 41at para 40.

[27] See s 213(3)(c) of the 1988 Act.

[28] [1998] FSR 63.

[29] See below, where consideration is given to the argument that statutory amendment to the Registered Designs Act 1949 extends registered design protection to aspects of a design consisting of shape or configuration where those aspects of the designs are purely functional in purpose.

[30] (1935) 53 RPC 139 at p 152.

[31] Eventually re-enacted in the 1949 Act.

> It remains, therefore, to consider whether the Design is new or original. Before attempting to answer this question it is necessary to determine which of the four features is sought to be protected. Is it shape, configuration, pattern or ornament? Shape and configuration are for all practical purposes considered as synonymous (see *Bayer's Design*, *ubi supra*, at p 80). Each signifies something in three dimensions; the form in which the article itself is fashioned. "Pattern" and "ornament" can, I think, in the majority of cases be treated as practically synonymous. It is something which is placed on an article for its decoration. It is something substantially in two as opposed to three dimensions. An article can exist without any pattern or ornament upon it, whereas it can have no existence at all apart from its shape or configuration.[32]

This was also the view of the learned authors of Laddie, Prescott and Vitoria (*The Modern Law of Copyright*)[33] which was cited with approval by Etherton J in *Lambretta*. However, Etherton J did not accept fully the view of the learned authors that Pumfrey J in the *Mackie* case had held that configuration is wider in concept than shape and covers features deriving their merit from the selection of the components used and the manner in which they are to be interconnected, rather than their physical layout. This would suggest that Pumfrey J was of the opinion that configuration referred to aspects of a design other than spatial arrangement.

Apart from the fact that Pumfrey J had been considering the rather special circumstances surrounding circuit diagrams, Etherton J could not agree with this conclusion. He noted that:[34]

> As I have said, that particular issue is not an issue with which I am concerned. There is no indication at all in the judgment of Pumfrey J that he was of the view that "configuration" in the statutory definition of "design" in PT III of the 1988 Act means anything other than three dimensional, rather than two dimensional, elements of an article,[35] and in that respect, bears the same meaning as in s1 of the 1949 Act, prior to the substitution of a new s1 by the 1988 Act.

Accordingly, Etherton J, regarded the Registered Designs Act 1949, section 1[36] and the interpretive case law on the meaning of configuration and shape relating to that provision as being of assistance in determining the nature of these concepts for the purposes of design right. In this context, he also regarded the matter as settled, in that configuration and shape, for the purposes of design right, referred to the relative arrangement of the elements of the design in *three* dimensions. It followed that the claimants had no design right under Part III of the 1988 Act in respect of the *two* dimensional colour patterns in their track top. His Lordship however, accepted that the claimant could enjoy artistic and literary copyright in aspects of their design documents.

The parties appealed. Nevertheless, before the Court of Appeal judgment is considered in this context, the consequences of Etherton J's first instance decision should be reviewed.

[32] Nevertheless, it should be noted that Luxmoore J made reference to pattern or ornament being aspects of a design which would be reproduced "substantially" in two dimensions. This emphasises the fact there is a practical and factual distinction between the aspects of a design in three dimensions from aspects of a design in two dimensions drawn by his Lordship and approved by Etherton J. This further suggests the distinction is a practical rule of thumb, that seeks to distinguish shape or configuration from pattern and ornament, but not a rule of law. This issue will be re-addressed below, where the Court of Appeal in *Lambretta*, *per* Jacob LJ, re-examined the utility and validity of this distinction as a means of determining the nature of shape and configuration as concepts within the design right régime.

[33] *Op cit*, see Vol 2 pp 2170–2171.

[34] [2003] RPC 41 at para 48.

[35] *ie*, represented in a design.

[36] As then drafted.

## THE RELATIONSHIP BETWEEN COPYRIGHT AND DESIGN RIGHT

Following the judgment of Etherton J in *Lambretta*, it is submitted that the combined effect of section 213 and section 51 of the 1988 Act is to preserve, albeit in an attenuated form, an overlap between copyright and design right in the protection of functional designs. Any aspect of a design, in order to attract protection under Part III of the 1988 Act, must, it is suggested, be capable of reproduction as an article in three dimensions.[37] These aspects of the design which are capable of protection must be functional in purpose. Nevertheless, aspects of the design, although protected by design right, may also be protected by copyright. Section 51(1) of the 1988 Act provides that:

> It is not an infringement of any copyright in a design document or model recording or embodying a design for anything other than an artistic work or a typeface to make an article to the design or to copy an article made to the design.

In his Lordship's opinion, the effect of section 51(1) was that, where a third party reproduced the aspects of a design protected by design right in three dimensions, that is, by way of the production of an article made to the design by an industrial process, then the owner of the design had a right of action restricted to infringement of design right, or possibly actions in respect of passing off or trade mark infringement. However there would not, in these circumstances, be any right of action based in copyright. Where the infringement took place other than by way of reproduction of an article in three dimensions, for example by unauthorised reproduction of the design document by photocopying the same, then an action in copyright would lie, in addition to any other forms of action.[38]

Accordingly, copyright still has a role to play in the protection of functional designs. The following practical situations may be considered in reviewing the continuing relationship between design right and copyright insofar as these régimes may co-exist in respect of a functional design. The following examples illustrate, as has been noted, that copyright still has a role to play in the protection of functional designs, albeit nowadays somewhat circumscribed.

## RELATIONSHIP BETWEEN DESIGN RIGHT AND COPYRIGHT

*Drawings that Constitute Artistic Works*

Where a document or drawing is produced, albeit not for the purposes of creating a blue print for the industrial manufacture of the subject matter of the document or drawing, then the work is a work attracting copyright protection. Even where there is no aspect of artistic quality in creating the work, be it a drawing, [39] it will attract copyright protection as a work of artistic copyright.[40] Where the work, for example in the form of a model, constitutes a work which has both artistic intent and aspects of

[37] But see a qualification to this view of the law in the case of *Lambretta* in the Court of Appeal, *per* Jacob LJ, which is considered below and which itself is subjected to critical analysis.

[38] This aspect of Etherton J's judgment was considered by the Court of Appeal and effectively endorsed. It is submitted that his Lordship set out an analysis of the effect and ambit of s 51(1) which is coherent and logical. The full effects of this aspect of his Lordship's judgment is discussed in Scanlan, "The Future of Design Right – Putting S51 Copyright Designs & Patents Act Act 1988 in its place" *op cit*, and will be considered only in outline in this article.

[39] Including works such as diagrams and graphic works.

[40] See section 4 of the 1988 Act which provides that: "... Artistic work means–
1(a) a graphic work, photograph, sculpture or collage, irrespective of artistic quality, ...
(c) a work of artistic craftsmanship.
A 'graphic work' includes
(2) (a) any painting, drawing, diagram, map, chart or plan".

craftsmanship, it will attract copyright protection as a work of artistic craftsmanship.[41] The unauthorised reproduction by a third party of such drawings or documents or models either in two or three dimensional forms will constitute an infringement of copyright in these works, whether or not the reproduction is within the context of an industrial application or process.[42] Section 51 is simply inapplicable to these cases.

## ARTISTIC WORKS SUBJECT TO AN INDUSTRIAL PROCESS BY COPYRIGHT OWNER OR LICENSEE

A work, *ie*, a drawing, which sets out the shape and configuration of an article, although it may be based on a work which is subject to copyright protection[43] may be created solely with the intent that it is to be a blueprint for the production by an industrial process of an article in three dimensions. The purpose of such drawings is to provide a means of exploiting commercially any artistic work created by the original drawings.[44] In these cases where the copyright owner or a licensee produces and markets articles from these "industrial" drawings, then any subsequent unauthorised reproduction by a third party of articles based on these *articles* is subject to the 1988 Act, section 52. This section provides that:

> ... where an artistic work has been exploited, by or with the licence of the copyright owner, by–
> (a) making by an industrial process articles falling to be treated for the purposes of this Part as copies of the work, and
> (b) marketing such articles in the United Kingdom or elsewhere.
>
> (2) After the end of the period of 25 years from the end of the calendar year in which such articles are first marketed, the work may be copied from by making articles of any description, or doing anything for the purpose of making articles of any description, and anything may be done in relation to articles so made, without infringing copyright in the work.
> (3) where only part of an artistic work is exploited as mentioned in subsection (1) subsection (2) applies only in relation to that part.

*The Effect Of Section 52*

Accordingly, section 52 provides that any unauthorised three dimensional reproduction by a third party of an article based on an artistic work, within the terms of the section, will constitute an infringement of the copyright in that work. However, the duration of the copyright in the original article produced and marketed by the copyright owner or licensee is limited to 25 years, the same maximum period afforded to designs subject to registered design protection.[45]

[41] Although what constitutes a work one of artistic craftsmanship is far from clear, see *Hensher Ltd v Restawhile Upholstery (Lancs) Ltd* [1975] RPC 31; *Merlet v Mothercare plc* [1986] RPC 15 and *Vermaat v Boncrest(No 1)* (2000) 97(25) LSG 40, *The Times* 23 June 2000. This issue is considered further below.

[42] See *King Features Syndicate Inc v O and M Kleeman Ltd* [1941] AC 417, [1941] 2 All ER 403.

[43] For example, a work which portrays a figure originally created as a cartoon character, such as Popeye, as part of an artistic or literary work.

[44] By the adaptation of the original drawings which created the relevant figure, so as to render the reproduction of the subject matter of the original drawings more readily amenable to an industrial process.

[45] See the Registered Designs Act 1949, s 8. The relationship between design right and registered design will be briefly considered below.

## CREATION OF A DESIGN DOCUMENT AS AN ORIGINAL WORK WHICH CONSTITUTES AN ADAPTATION OF AN EARLIER ARTISTIC WORK BY THE COPYRIGHT OWNER

It should be noted that section 52 refers to an artistic work *having been* exploited. Thus section 52 is inapplicable where a copyright owner produces drawings (which may be defined as industrial drawings), intended to form the blueprint for the industrial production of an article, albeit based on an earlier work,[46] be it a drawing, model or sculpture. In this case the following situations may arise.

Where a third party without authorisation makes articles based on the "industrial" drawings, it is suggested that that there is an infringement in any design right that exists in the drawings and that section 51 is applicable in these cases. Where the third party copies either exactly or substantially the industrial drawings in two dimensions, *eg* by photocopying these drawings, section 51 is excluded, and the unauthorised copying is, according to Etherton J in *Lambretta*, both an infringement of the copyright in the drawings and an infringement of any design right. An example of such copying which may occur on a comparatively regular basis may be the reproduction of the figures the subject of the industrial drawings or model as a two dimensional figure on a tee-shirt.[47]

In the above cases, any reproduction of the *original* drawings, which formed the basis of the industrial drawings, by a third party, be it by way of two or three dimensional reproduction, and whether or not by virtue of an industrial process, will be an infringement of the copyright in those drawings. Section 51 is, of course, inapplicable in these cases.

### *The BBC Worldwide Case*

A case that may be regarded as relevant to the above discussion is *BBC Worldwide Ltd and Another v Pally Screen Printing Ltd and others.*[48] Although the case concerned an application for summary judgment,[49] Laddie J set out, albeit on a provisional basis, his opinion on the relationship between copyright in designs and any subsisting design right, in part based on an interpretation of the effect and application of the 1988 Act, section 51.

### *The Facts Of The Case*

The facts of the case, in outline, are that the defendants collectively were responsible for reproducing, inter alia, unauthorised artwork on certain garments, namely, the reproduction of the famous Teletubby figures on articles, including tee-shirts. The plaintiffs who owned the intellectual property rights in the figures[50] were nevertheless unable to identify precisely the works[51] alleged to be infringed. That is to say, they could not establish whether the unauthorised copying had taken place from photographs of the figures set out in an edition of a BBC publication, or directly from other artistic work of the plaintiffs which portrayed the figures.

[46] Which enjoys copyright protection either as an artistic work or a work of artistic craftsmanship.

[47] In these cases the industrial drawing of a figure may have been created with simplicity of shape, as compared to the figure in any original drawings, on the basis that such simplification may aid manufacture. This very advantage may also aid the reproduction of the figure by the third party in the circumstances set out in the main body of the text above.

[48] [1998] FSR 665.

[49] Under Ord 14, Rules of the Supreme Court, see now Civil Procedure Rules 1998, Part 24 and PD 24.

[50] Although this issue was not entirely free from doubt, the summary nature of the action and the decision of Laddie J in refusing to grant summary judgment rendered this issue somewhat academic.

[51] Which set out and created or reproduced the figures of the Teletubbies.

Although his Lordship dismissed the plaintiffs' application for summary judgment he made specific reference to the possible application of section 51 to the case. In the words of his Lordship:[52]

> The defendants argue that if the design is for an article which is not itself an artistic work than s51 comes into operation. It is not said before me that the Teletubby characters – that is to say, the three dimensional puppets – are artistic works or unarguably artistic works. So, since they are arguably not artistic works the defendants say that there are two categories of activity which are excluded from being treated as infringements of copyright under the provisions of s 51(1). The first consists of making articles to the design. So, they say that since the designs here were for non-artistic three dimensional works or articles it is not an infringement of the copyright in those design documents to make *articles* to the design.

In part his Lordship's view that the three dimensional figures were not unambiguously artistic works is based on a line of authority that a three dimensional figure can only constitute a sculpture and therefore an artistic work where the work is the work of a sculptor.[53] The problem with this line of authority is that it is difficult to apply this principle with any degree of certainty,[54] with consequential difficulties as to the nature of protection to be afforded to the work in cases of infringement. It is suggested by way of contrast to Laddie J in the *BBC Worldwide* case that the approach of Etherton J in the *Lambretta* case is to regard a design document or model as being a work capable of being protected as a work of copyright[55] and/or as a design. This, it is suggested, is the more logical and practically effective approach to the interpretation of the nature of the work, and the consequential form of protection to be ascribed to the work in cases of infringement. It is suggested that the approach of Etherton J concentrates not on the intention (presumed or purported) on the part of the creator of a work as to the nature of the work, but on the nature of the infringing act. The consequences of the law adopting this reasoning and the effect upon the 1988 Act, section 51 will be considered in more detail below and in the conclusion of this article. At this stage it can be said that such an approach promotes certainty of application of section 51 in cases involving infringement of a work, and removes any difficulties in determining whether a design document or model can also constitute a work of copyright in addition to being a work which is subject to design right.

## THE NATURE OF INFRINGEMENT IN THE *BBC WORLDWIDE* CASE: COPYING AN ARTICLE MADE TO A DESIGN

The question whether the copying of the Teletubby figures was three dimensional was not in issue in any event in the *BBC Worldwide* case. If it had been, it is suggested that

[52] [1998] FSR 665 at p 671.

[53] See Stokes "Categorising Art in Copyright Law" [2001] Ent LR 179 and the cases cited therein, particularly *J & S Davis (Holdings) Limited v Wright Health Group Limited* [1988] RPC 403; *Metix (UK) Limited and Another v G H Maughen (Plastics) Limited and Another* [1997] FSR 718, *per* Laddie J.

[54] This approach to the ascribing of copyright protection to artistic works is even more problematic when applied to that special form of artistic work: a work of artistic craftsmanship. Despite the judgment of Walton J in *Merlet v Mothercare plc* [1986] RPC 15, following the House of Lords' decision in *George Hensher Ltd v Restawhile Upholstery (Lancs) Ltd* [1975] RPC 31, whether a work constitutes a work of artistic craftsmanship remains unclear. Is it the intent of the creator which is the determining factor, or the perception of the public as to the nature of a particular work? In the more recent case of *Vermaat v Boncrest (No 1)* (2000) 97(25) LSG 40, *The Times* 23 June 2000, the court observed that there being no clear definition as to what constituted "artistic craftsmanship", it concluded that it was necessary to determine whether the work in question could fairly be said to be a work of both a craftsman and an artist. This is hardly clear, and is an unsatisfactory approach to the question of determining the nature of a work for the purposes of ascribing the form of the protection afforded to the work from unauthorised reproduction by third parties. Space and the balance of the article precludes further consideration of this issue.

[55] Which must be based on the work being an artistic work or a work of artistic craftsmanship.

there would have been no doubt that section 51 could have applied and the infringing act could have been founded in design right, and not copyright. However, Laddie J addressed the question that was central to his judgment. His Lordship noted that counsel for the defendants had argued that:[56] ". . . section 51(1) provides that it is not an infringement to copy an article made to the design."

It followed from the above, in defendant counsel's submission in the case, that:[57]

> Any activity which constitutes copying an article – that is to say, copying a three dimensional article made to the design . . . cannot constitute an infringement.[58] So, Mr Fitzgerald (counsel for the defendants) says, here it is possible, and indeed likely, that the artwork of which complaint is made was copied from the Teletubby puppets seen on television. They are indirectly copies of the puppets, and if that is so they cannot infringe any design in the design documents – that is to say, the plaintiffs' drawings.

Counsel for the plaintiffs, Mr Baldwin QC, argued that this view as to the effect of section 51 could not be right. Both counsel could agree that the unauthorised making by third parties of three dimensional articles from any design was an infringement of the design right in either a drawing or model.

However, again in the words of Laddie J,[59] counsel for the plaintiffs had argued that: ". . . the words 'or to copy an article made to the design' should be construed as if they read as follows, 'To copy an article made to the design and so as to make something which is a three dimensional article'".

Although his Lordship could not agree with this proposition, it is suggested that this is precisely the position the law may take in certain circumstances following the *Lambretta* case.[60]

## THE RELATIONSHIP BETWEEN THE *LAMBRETTA* CASE AND THE *BBC WORLDWIDE* CASE

If the effect of the *Lambretta* case is that a design document or model incorporating a design are each capable of being both an artistic work and a work attracting design protection, then whether the unauthorised copying of the work constitutes an infringement of a design right or of copyright (or possibly both), depends not on the intention of the party creating the work, but on the nature of the infringement. Where the infringing act takes the form of a three dimensional article, whether based on the copying from a two dimensional design document or a three dimensional model, then section 51 applies and the act is to be regarded as an infringement of the design right in the original works, but that there is no infringement of the copyright in the work.[61]

Where the infringing act takes the form of the copying of a design document or any model incorporating or embodying a design by way of two dimensional reproduction,

[56] [1998] FSR 665 at p 671.

[57] In the words of Laddie J, *ibid.*

[58] That is to say of any copyright in the design.

[59] *Ibid*, at p 672.

[60] Subject to the view of Jacob LJ in *Lambretta* in the Court of Appeal which is considered below.

[61] It is suggested, for reasons set out above and below, that this statement remains the law for all practical purposes, even after the view of Jacob LJ in *Lambretta* in the Court of Appeal as to the utility and validity of the three dimension concept as a means of determining the nature of design right and the ways in which infringement of the right may take place.

then it is suggested, following Etherton J in *Lambretta*, that section 51 is inapplicable and the act is an infringement of the copyright in the original work, be it a design document or a model.[62]

If this analysis of the nature and effect of section 51 is applied to the reproduction of figures such as in the case of the Teletubbies, the following conclusions can be drawn. The reproduction of such figures by way of three dimensions is, in essence, an industrial exploitation of the work, and therefore the copying must be regarded as an infringement within the context of design right.[63]

Where, as in the actual *BBC Worldwide* case, the infringing act takes the form of a two dimensional copying of a figure either from a two dimensional design document or a three dimensional model, the copying is, it is suggested, principally the infringement of the copyright which exists in the original work the subject of the infringing act and is not an industrial exploitation. Again, using the facts of the *BBC Worldwide* case, the industrial production took place in respect of the manufacture of the tee-shirts, the reproduction of the Teletubby figures by way of a two dimensional print upon the tee-shirts was clearly an infringement of the copyright subsisting in the original works.[64] Where the copying is of an *article*[65] which has already been manufactured or marketed in the United Kingdom or elsewhere by the copyright owner or a licensee, then such copying is subject to the 1988 Act, section 52, the nature and effect of which has been considered above.

This approach to the nature of infringement that may take place with respect to design documents and models that are adaptations of already existing fictional characters,[66] originally portrayed in literary and dramatic works, but which are to be the subject of manufacture in three dimensional forms is, it is submitted following *Lambretta*, both simple and capable of certain and sure application. Both the creators of the relevant works and infringers would be aware of the consequences of infringement in these situations.

## INDUSTRIAL DESIGNS BASED ON EXISTING FICTIONAL CHARACTERS

Where an original artistic work, such as a cartoon figure, forms the basis for a an industrial design created in a design document or a model, a prime example being an industrial article which includes aspects of a figure such as Mickey Mouse in a so-called Mickey Mouse telephone, the consequences of unauthorised reproduction by third parties of an article of this kind are broadly similar to the copying of figures such as the Teletubbies and other fictional and novelty characters, and which has been noted above. In the case where the character forms only an aspect of the design, as in the case of the Mickey Mouse telephone, then, since the drawing or model incorporating

[62] And as has been noted above a possible infringement of the design right in the relevant drawing or model. See the possible qualification to this view of the law noted in fn 60 and fn 61 above.

[63] Furthermore, as Etherton J noted in the *Lambretta* case ([2003] RPC 41 at para 69 of the report) the copying of the designs by way of three dimensional articles may also constitute the tort of passing off, or breaches of intellectual property rights other than design right, and excluding copyright.

[64] Although Etherton J in *Lambretta* also suggested that the infringement in such cases could also constitute an infringement of any design right in the drawings or model, this may not have a significant practical effect however, since both the nature of the protection afforded in cases of design right and its duration as compared to that which exists in respect of copyright may render the supplementary protection under design right of little significance.

[65] Based on an artistic work.

[66] Or figures which are not created in such circumstances as part of a character merchandising operation but are created with the express intent of the ultimate production of model figures for example as toys or novelty items. These works, it is suggested, would be subject in cases of infringement to the same forms of protection noted in the main body of the text.

the design is produced as part of a process that is intended to end in the mass production of a utilitarian article, the drawing incorporating the design is clearly a design document, and is subject to design right protection, as is any model incorporating the design. In these cases the unauthorised reproduction by a third party of an article in three dimensions based on these drawings, or any model, must be an infringement of the design right in the drawings or model, and section 51 would be applicable. The two dimensional copying of the drawings or the model incorporating the design would, as in the cases above, constitute an infringement of the copyright in the drawings or the model.[67] However, in these cases the mere copying of the drawings would be an unlikely event in isolation, since the only effective purpose in copying the drawings would ultimately be to reproduce on an industrial basis an article in three dimensions based on such drawings.

## DESIGN RIGHT AND REGISTERED DESIGN RIGHT SUBSISTING IN DESIGNS

Of course, the question whether aspects of a design document incorporating an essentially functional design such as a "Mickey Mouse" telephone also include aspects of the design which are capable of protection under the Registered Designs Act 1949, must also be considered. Such designs may be protected insofar as the shape configuration, pattern or ornament constituting the design have aesthetic elements, at least this was regarded as the law prior to the implementation of the Design Directive.[68] Accordingly, whether designs such as novelty telephones, incorporating aspects of a pre-existing fictional character are capable, in addition to being protected by design right, and copyright, of being also, in part, capable of being protected by registered design needs to be considered.[69]

## REGISTERED DESIGN

*Registered Design And Design Right*

A design document may include aspects of a design that are functional in purpose and therefore to be protected by design right. Such designs may also have aspects which are aesthetic in purpose and appeal to the eye. In these circumstances, these aspects of the design may be protected by registered design right under the Registered Designs Act 1949. The following examples may help to illustrate this principle.

Designs made in respect of toy figures set out in design documents that are intended to form the blueprint for the industrial production of such figures are, irrespective of the appeal of such figures, essentially functional in nature. Even where the figures are based on already existing fictional characters, in which the individual creator may enjoy artistic copyright, the designs will generally have regard to ease and cost of production, as well as safety aspects so as to ensure that the figures are safe to use by children. Designs of this nature are therefore, for the reasons set out above, essentially functional in nature and therefore principally protected by design right.

[67] And it must be noted that, according to Etherton J in *Lambretta*, an infringement of the design right existing in any drawings or models.

[68] See the Registered Designs Regulations 2001 SI No 3949 and below.

[69] This issue was partially discussed in Scanlan and Gale "Industrial Design and the Design Directive: Continuing and Future Problems in Design Rights?" *op cit* at p 103–109.

If, by way of contrast, consideration is given to a design for an item of jewellery, differences in the ways in which the law regards such a design from the example set out above may be noted. An important aspect of such an item is its appeal to the eye. In these circumstances the design is principally to be protected by recourse to the Registered Designs Act 1949. This is certainly the case in respect of the aspects of the design relating to pattern and ornament, which are aspects of a design set out in a design document that are not capable of being protected by design right, but any aspects of the design that consist of shape or configuration may, if they appeal to the eye, be protected under the registered design régime.

As a final example, consideration may be given to designs of "novelty items" such as the Mickey Mouse telephone, as has been noted above. The design in these circumstances is essentially functional in nature and, therefore, those aspects of the design possessing this attribute and which consist of shape and configuration may be protected by design right. However, the appeal of this particular type of telephone in part may be due to the fact that it is a novelty item reproducing aspects of a celebrated fictional character. The design in these circumstances may constitute aspects of shape or configuration that possess the attributes of eye appeal, and may therefore, to this extent, be protected as a registered design.[70]

## THE ROLE AND FUNCTION OF REGISTERED DESIGN

Although the *Lambretta* case did not consider the role and function of registered design, it is suggested, for the reasons set out above, that a consideration of this intellectual property right is appropriate, at least in outline. Prior to the amendments made by the Registered Designs Regulations 2001[71] (the Regulations) to the Registered Designs Act 1949 (the 1949 Act), it was determined by authority that the registered design régime, although applying to designs which included both shape and surface decoration,[72] was nevertheless restricted to these aspects of designs only insofar as they had eye appeal,[73] whereas design right within the terms of Part III of the 1988 Act could only subsist in aspects of shape and configuration in designs that were functional in purpose. It may well be, however, that the new section 1C inserted into the 1949 Act by the Regulations, when combined with the new section 1(3), provides that the aspects of designs consisting of shape or configuration and which are essentially functional in purpose may nevertheless now be capable of being registered under the 1949 Act.

### *Aesthetic And Functional Designs – The New Relationship*

As a result of the new section 1(3) of the 1949 Act, inserted into the Act by the Regulations, the registered design régime is capable of applying to designs in respect of any industrial or handicraft item. However, by virtue of the new section 1C(1), no registered design right can subsist in "features of appearance of a product[74] which are *solely*[75] dictated by the product's *technical*[76] function."

[70] Those aspects of the design which consist of pattern or ornament can of course only be protected under the registered design régime.

[71] SI No 3949, which came into effect on 9 December 2001 and which implemented the Design Directive 98/71 EC OJ 289, 28/10/1998 p 0028–0035.

[72] By way of shape, configuration and pattern or ornament.

[73] See *Amp Inc v Utilux Pty Ltd* [1972] RPC 103.

[74] S 1(2) provides that a design for the purposes of the 1949 Act "means the appearance of the whole or a part of a product resulting from the features of, in particular, the lines, contours, colours, shape, texture or materials of the product or its ornamentation".

[75] Italics supplied.

In the absence of case law it is difficult to determine the ambit and nature of this latter provision, the purpose of which is to exclude certain designs from protection under the registered design régime. However, it is suggested that the combined effect of the above provisions leaves open the possibility that only where aspects of a design[77] can take the adopted form and no other, will a design be regarded as dictated solely by its technical function, within the terms of section 1(C) and therefore excluded from the ambit of the 1949 Act. Any design, which could adopt an alternative form than the actual form adopted, and still function as an effective design,[78] may not be a design dictated solely by its technical function within the terms of section 1C, and could therefore be registered as a design under the provisions of the 1949 Act, despite the fact that the design is essentially functional in nature and purpose. This would include a large number of designs, the aspects of which prior to the amendments to the 1949 Act, as noted above, could only be protected as a design under Part III of the 1988 Act. If this interpretation of the nature of the amended 1949 Act is accepted, many functional designs would be subject to the provisions of the 1949 Act, with a consequential reduction in the importance of the design right régime prescribed under Part III of the Copyright Designs and Patents Act 1988 and a consequential reduction in the importance of section 51 of that Act.[79]

## THE HANDICRAFT ITEM ISSUE

It should be noted that the definition of a design for the purposes of the amended 1949 Act is connected with the concept of a product. A product is defined by section 1(3) of the Act as including any industrial item or *handicraft item*. This latter term could be interpreted as applying to works and products, which under the pre-Directive law were regarded exclusively as works of artistic craftsmanship. If this interpretation of the effect and nature of the term is subsequently adopted by the courts, works of artistic craftsmanship set out initially in drawings could be subject to registration under the 1949 Act. This would further reduce the possible role of design right and the possible application of the 1988 Act, section 51 where infringement took place in respect of such works.

However, an alternative interpretation of the nature and effect of the term "handicraft item" is that only designs which "mimic" a work of artistic craftsmanship, but which is a design in which the article set out in the design is intended to be subject to an industrial process will be subject to the provisions of the 1949 Act. The nature of these kinds of product is that they invoke nostalgia in purchasers and users, and are intended to reproduce in the finished manufactured article the idea of the "Old English Artisan", and the hand crafted domestic products of some ideal English past. Usually these products boast that they are in some way and in some degree made by hand, *eg*, they may, after production by an industrial process, be hand painted, nevertheless they are for all practical purposes mass produced articles.[80]

If the term handicraft item were so interpreted, and applied to designs so as to include within its ambit only designs in respect of the kind of products noted above,

[76] Italics supplied.

[77] Which consist of shape or configuration.

[78] Albeit that the alternative design might be less efficient or more expensive to manufacture.

[79] For a detailed discussion of the effect of the implementation of the Design Directive and corresponding amendment to the 1949 Act, see Scanlan and Gale "Industrial Design & The Design Directive: Continuing and Future Problems in Design Rights" *op cit.*

[80] An example of these kinds of goods and articles are those sold in shopping chains such as Past Times.

the new provisions of the 1949 Act would not, at least in this context, expand the ambit of registered design. The creation by a party of a drawing or model meant to be a work of artistic craftsmanship would be protected principally by copyright, and the reproduction of the same by a third party, in either three or two dimensions, would be an infringement of copyright in the work.

Where the owner of the copyright or a licensee of such a work manufactured the product represented in the drawing by an industrial process and marketed the same in the United Kingdom or elsewhere, then the 1988 Act, section 52 could apply in cases of any infringement.

However, where a drawing or model, albeit potentially a work of artistic craftsmanship, was nevertheless intended to form the blueprint for the mass production by an industrial process of the relevant article set out in the drawing or model, then it is suggested the drawing or model would constitute a handicraft item, within the terms of the 1949 Act, section 1(3). In these circumstances the principal means of protection from the unauthorised reproduction of a three dimensional article based on the drawing or model would be by virtue of the provisions of the 1949 Registered Designs Act.

## THE COURT OF APPEAL IN *LAMBRETTA*: SHAPE, CONFIGURATION AND THE THREE DIMENSION/TWO DIMENSION DISTINCTION

The Court of Appeal in *Lambretta* was principally concerned with examining the nature of the test of originality to be applied for the purposes of determining if design right could exist in respect of aspects of a design contained in a design document or a model, together with the nature of infringement of design right. These issues for the reasons of balance of the article and space are not considered in this article. However, the Court of Appeal[81] were in broad agreement as to Etherton J's opinion at first instance both as to the nature and meaning of shape and configuration as aspects of a design within the design right régime and the analysis of the nature of design right, together with the role and function of the 1988 Act, section 51. Although Jacob LJ was critical of the suggestion that configuration[82] can only exist in three dimensions, his Lordship noted that:[83]

> Before I consider further Lambretta's contentions[84] on this point. I should get out of the way what I consider to be a peripheral dead-end. This is the question of two or three dimensions. It will be noted that Luxmoore J in part reasoned that shape and configuration were three-dimensional. Before Etherton J in this case there was debate about this too (see paragraphs 33–49). Etherton J concluded that "configuration" in the meaning of "design" in Part III of the 1988 Act refers to the relative arrangement of three-dimensional elements.
>
> I do not think that a debate about dimensions assists. All articles (even thin flat ones) are 3 dimensional (using the practical Euclidean view of the world – not that of modern day physics). There is no reason why a "design" should not subsist in what people would ordinarily call a "flat" or "2-dimensional" thing – for instance a new design of doily would have a new "shape" and could in principle have UDR in it.
>
> Thus dimensions do not really come into it. The real question is whether the words "configuration ... of an article or part of an article" can cover the mere colouration of

[81] See in particular Jacob and Sedley LJJ.

[82] And presumably shape as the terms are regarded as synonymous.

[83] [2004] EWCA Civ 886 at para 23 *et seq.*

[84] On the nature and meaning of shape and configuration for the purposes of design right.

> parts of an article. Mr Wyand [85] submitted they could – that it would not be an abuse of language to say that the red, blue and white components of the Lambretta top[86] were "configured" together. That is true but beguilingly simple – the truth is that the components are indeed configured to produce the ultimate complete article but their colour has nothing to do with that configuration.

*The Three Dimension/Two Dimension Question*

Jacob LJ's comments with regard to the "three-two dimension question" should be subjected to a critical review. This part of his Lordship's judgment must be considered as comments within the context of a judgment concerned principally with other issues. Furthermore, these comments are not supported specifically by his brother judges, and are in potential conflict with a preponderance of earlier authority. In addition, the practical consequences of this opinion need to be discussed.

Although Jacob LJ seems to be entirely dismissive of what may be defined as the two or three dimension test, as a means of differentiating which aspects of a design may be subject to protection under design right, it is suggested that this analysis or approach to determining the nature and meaning of shape and configuration has some utility, at least as regards the application of section 51 to functional designs. If we accept his Lordship's approach or opinion that all designs and articles, when reproduced, are in three dimensions, it may be argued that this merely recognises principles of geometry and physics.[87] If, however, we take this as fact, but nevertheless regard the previous view as expounded by Etherton J at first instance, that shape and configuration relate to aspects of designs, intended to be produced in finished articles based on the design in three dimensions, not as a legal requirement,[88] but as a practical rule of thumb that can be applied in the majority of cases so as to give guidance whether aspects of a design are capable, *prima facie*, of protection under Part III of the 1988 Act, it is suggested the law has a workable and practical formula in determining the ambit and nature of design right. That there will be borderline cases where this rule of thumb does not work effectively does not invalidate this "test" or analysis of the nature of shape and configuration for the purposes of determining whether aspects of a design are capable of protection under Part III of the 1988 Act.

## THE THREE/TWO DIMENSION QUESTION AND SECTION 51

Furthermore, this established analysis or test for determining the nature of shape and configuration within functional designs is, it is suggested, particularly useful when considering the application of the 1988 Act, section 51 in respect of such designs. This can be illustrated if we take the example of Jacob LJ when considering a design in respect of a doily. The doily, when represented in the design document, is effectively in two dimensions. When the article is manufactured it clearly exists in three dimensions.[89] Applying Etherton J's approach to the application of section 51 to such a design and article, if a third party reproduced the doily without authorisation, that would be an industrial reproduction of the aspects of the shape or configuration of an

[85] Counsel for the claimant.

[86] *ie* the colourways of the Lambretta track top.

[87] As his Lordship accepted.

[88] And as a fact of existence.

[89] As may any aspects of pattern and ornament, which if embossed would also be reproduced in three dimensions, nevertheless these aspects of the design would irrespective of their dimensions be protected under the Registered Designs Act 1949 as aspects of a design which are aesthetic in nature and appeal.

article made to a design in three dimensions. In such a case section 51 would apply, and the reproduction to this extent would be an infringement of the design right in the article, but there would be no infringement of copyright.

Where the infringement takes the form of copying the whole design document, for example the unauthorised reproduction of the design document in an article or a "trade magazine", then the reproduction, notwithstanding Jacob LJ's view that all objects have three dimensions, is, in effect, a reproduction in two dimensions, and accordingly not a reproduction of the design by way of reproduction of an article, which it is suggested must be a reproduction in three dimensions. In these circumstances, section 51 does not apply, and the reproduction is both an infringement of the design right in the design document, and an infringement of any copyright. In effect, the reproduction in these circumstances is the same as the reproduction of the colourways of the claimant's track top in the *Lambretta* case. The reproduction of the design of the doily in an article or trade journal is effectively by means of ink impregnated in the paper, and the reproduction must therefore, as in the case of the colourways in the claimant's track top in *Lambretta*, be regarded as being in two dimensions. It is for these reasons submitted that Jacob LJ's views as to the validity or utility of the two-three dimension test within the context of the design right régime, although correct in principle and strict logic, may, nevertheless, have little if any practical import when regard is had to the application of the 1988 Act, section 51 to functional designs.

The Court of Appeal, in rejecting, as did Etherton J at first instance, any subsisting design right in the claimant's track top as regards the colourways, nevertheless considered the possible future application of the new Community Design Regulation and the new community design right to designs like those of the Lambretta track top. The Court of Appeal was, therefore, not entirely negative in its consideration of situations where the law of intellectual property may provide protection in respect of designs which are similar to that of the claimant's in *Lambretta*, at least for future designs and possible future claimants. The nature of this new community design right should therefore be considered, albeit briefly.

## THE COMMUNITY DESIGN RIGHT RÉGIME

Jacob LJ in his judgment in *Lambretta* had noted in passing that:[90]

> I leave the appeal with one further observation: that the new European unregistered design right,[91] although lasting a shorter period than the UK UDR,[92] clearly would cover this case. Art 3(a) of Regulation 6/2002 says:
>
> "design means the appearance of the whole or a part of a product resulting from the features of, in particular, the lines, contours, *colours*, [93] shape, texture and/or materials of the product itself and/or its ornamentation"
>
> This Regulation was not in force at the time when the events of this case occurred. But it demonstrates how, if one wanted to cover a case such as the present, it could be done. Whether the UK still needs its own unregistered design right on top of that provided by the Regulation is perhaps open to question.

[90] [2004] EWCA Civ 886 at para 41.

[91] See Council Regulation EC No 6/2002 of 12 December 2001 on Community Designs.

[92] The period of protection is three years from the date on which the design was first made available to the public within the Community, see art 11 of the Regulation.

[93] Emphasis added.

A critical evaluation of his Lordship's comments on the role and function and relationship between the national and European design right requires its own separate article. Nevertheless, the following observations may briefly be made here. It is clear from the citation of article 3(a) of the Regulation, and from the opinion of his Lordship on the nature of that provision, that the design right created by the Regulation, although bearing in part a resemblance to the national design right under Part III of the 1988 Act, is much wider in the nature or forms or aspects of a design that it is capable of protecting than the national design right. It should also be noted that if a designer desires a period of protection greater than three years, then the designer must have resort to the provisions of Part III of the 1988 Act.[94]

Furthermore, since the Court of Appeal in *Lambretta* did not consider the relationship *per se* between design right and registered design, since this was not an issue in the case, it is suggested that the arguments set out above as to the nature of design right and the relationship between the various means by which the law may protect essentially functional designs, including the possibility that certain, if not the majority of, functional designs may now be capable of protection under the provisions of the Registered Designs Act 1949 as amended, are based on a presently uncontested interpretation of statute. In these circumstances it may be questioned whether we need an unregistered design régime at all, be it national or European.

## CONCLUSIONS

The judgment of Etherton J in the *Lambretta* case has provided much needed clarity as to the nature and effect of the 1988 Act, section 51 and the relationship between design right and copyright in the protection of functional designs. The advantage of Etherton J's approach to this issue is that it regards the provision as being essentially simple in its intent and purpose, and takes a literal and rather elegant approach, limiting the provision to its strict terms, and not seeking to import complexities as to the ambit and operation of the provision so beloved of academic commentators; an approach which has resulted in confusion and lack of clarity as to the relationship and demarcation between copyright and design right. It is therefore to be welcomed that the Court of Appeal endorsed in all material particulars this most important aspect of his Lordship's first instance judgment. As the law now stands following *Lambretta*, whether section 51 applies to a case of infringement of aspects of a design or model incorporating a design is, it is submitted, dependent, not on the presumed or purported intent of the party who created the work the subject of infringement,[95] but rather on the nature of the infringement that has taken place.[96] Where the infringement of a work, be it as regards aspects of a design in a design document, or a model incorporating a design, takes place within the context of an industrial exploitation, in essence by means of a manufactured article,[97] then infringement is restricted to

[94] Whether essentially functional designs should have a period of protection as great as fifteen years as provided at present under Part III of the 1988 Act is of course one of the issues that needs to be considered in view of the new Community design right. Furthermore, consideration needs to be given to the nature and ambit of the registered design régime created by the Regulation, see art 12 of the Regulation.

[95] *ie* is the work intended to be an artistic work or not.

[96] Although, as has been noted above, the exact nature of infringement of design right as considered in *Lambretta* is not discussed in this article.

[97] And for reasons stated above, a reproduction in three dimensions, despite Jacob LJ's opinion that this aspect of the reproduction is not conclusive or even helpful in analysing the nature of the aspects of the design, namely shape and configuration protected under the design régime created by Part III of the 1988 Act, see above.

infringement of any subsisting design right.[98] In all other cases infringement by way of copying aspects of a design, but outside what may be regarded as an industrial exploitation, is capable of being an infringement both of the subsisting design right and any co-existing copyright right. The effect of the *Lambretta* case[99] might not be to bring about the demise of section 51, but it has at least put the section in its place and context, with consequences that all intellectual property lawyers, both academic and practitioner, should welcome. Nevertheless, as the court at both first instance and in the Court of Appeal in *Lambretta* recognised, the law relating to the protection of functional designs is fraught with conceptual difficulties if not contradictions and conflicts. Some of these issues have been considered in this article. Nevertheless, it remains to be seen whether a comprehensive rationalisation and simplification of the law of intellectual property as it applies to the protection of designs in general is needed and, in particular, with regard to the supposed clear cut distinction between the means of protection afforded to aesthetic and functional designs, which, it is suggested, has possibly been eroded by recent amendments to the Registered Designs Act 1949.[100] In these circumstances it remains to be seen whether there is a need for either a national or a European unregistered design right. It may also be desirable to consider whether copyright has a role to play in any modern design régime, particularly where designs are produced in the modern world with mass manufacture in mind. These issues are, of course, beyond the power of the judiciary to implement, and will require legislation, not only on a national basis but on a European scale.

[98] And possibly other intellectual property rights, but excluding copyright.

[99] Particularly when taken in conjunction with the recent developments in the realm and function of registered design right which has been considered above.

[100] See above and Scanlan and Gale "Industrial Design and the Design Directive: Continuing and Future Problems in Design Rights" *op cit.*

NOTTINGHAM LAW JOURNAL

VOL 14(2) 2005

# CASE AND COMMENT

*The address for the submission of material for this section is given at the beginning of this issue*

## DISCLOSURE OF EXPERT EVIDENCE

*Nicos Varnavas Hajigeorgiou v Vassos Michael Vasiliou* [2005] EWCA Civ 236; [2005] 3 All ER 17 (CA)

(Brooke LJ, Dyson LJ, Gage LJ)

The judgment in the case of *Nicos Varnavas Hajigeorgiou v Vassos Michael Vasiliou* provides useful guidance to practitioners in the handling of expert reports.

### ISSUES

The case involved an alleged breach of the defendant's covenant of quiet enjoyment in a lease of premises in London. At a case management conference, the parties were given permission to instruct one expert each in the disciplines of restaurant valuation and profitability.

Prior to the hearing, the solicitor for the defendant submitted a witness statement to the court which, *inter alia*, proposed that one Mr W would be a suitable expert who could produce evidence on valuation and profitability. After the hearing, counsel for both parties drew up an order which they agreed reflected both what the judge had ordered and the agreed directions that had been approved by the judge. The order provided: "Both parties do have permission, if so advised, to instruct one expert each in the specialism of restaurant valuation and profitability".

Mr W inspected the claimant's premises. Subsequently, the defendant sought facilities for a second expert, Mr N, to inspect the same premises. As no reason was given by the defendant's solicitor (despite requests), the claimant's litigation manager telephoned Mr W who, it is reported, confirmed that he had already prepared his evidence.

The issue came before the court. At first instance, the judge ordered that:

> (a) the original case management order gave permission to the defendant only to rely on the expert evidence of Mr W;
> (b) the defendant needed permission from the court to rely on the second report of Mr N; and
> (c) the defendant could have permission to instruct Mr N, but only on condition that he disclose Mr W's report to the claimant.

The defendant appealed on two grounds:

> (a) the original case management order gave permission to both parties to instruct an expert in restaurant valuation and profitability. It did not give permission to the defendant only to instruct Mr W; and

(b) in any event, the judge whose order was appealed should not have made disclosure of Mr W's report a condition of granting permission for Mr N to be called as an expert witness instead.

The appeal was decided in favour of the defendant on the first ground. The defendant did not need the permission of the court to rely on the evidence of Mr N; and even if he did need permission, the judge was right to order, as a condition for granting this, the disclosure of Mr W's report.

## ANALYSIS

*Permission to rely on another expert*

The original order did not name Mr W. If it had, the defendant would have required the permission of the court to substitute Mr N for Mr W.

The difficulty here was that the defendant had prepared a witness statement before the first case management conference explaining that Mr W was a suitable expert to be instructed by him and in this sense, it was argued that the order was construed as giving the defendant permission to rely on the evidence of Mr W. However, the Court of Appeal decided that evidence in the witness statement pertaining to Mr W's suitability and charging rates was no more than evidence that there were experts in the field, that Mr W's rates were reasonable and proportionate and that, if the claimant wished to instruct a like expert, his fees were equally likely to be proportionate.

It was also pointed out that the court did not have the power to give permission for the *instruction* of experts, in accordance with CPR r 35.4. Rather, the court had the power to *restrict* expert evidence. On this basis, the defendant ought to have been able to instruct Mr N without seeking the permission of the court.

*Disclosure of the first report of Mr W*

This was dealt with as part of the appeal hearing, although the issue did not now strictly arise because the Court of Appeal had found in favour of the defendant on the first issue above.

Naturally, it was argued by the defendant that the undisclosed report of Mr W was privileged from production. Privilege had not been waived. It was submitted that the order now appealed effectively required the defendant to disclose Mr W's report and deprived him of his right to privilege over it.

In terms of the second issue, which the Court of Appeal approached on the basis that the defendant required the permission of the court to rely on a second expert, it was bound by the decision in the case of *Beck v Ministry of Defence*.[1] What the Court of Appeal drew from this case was that a court has the power to grant permission to a party to rely upon further expert evidence which it should usually do on the basis that the report of the first expert is disclosed.

The Court of Appeal highlighted the importance of the *Beck* case in showing how the court will control the way in which litigation generally, and expert evidence specifically, is managed. If a party needs the permission of the court to rely upon the evidence of a second expert in place of the first, and to avoid so-called expert shopping, that party will be required, first, to waive privilege over the first report.

[1] [2003] EWCA Civ 1043, [2003] CP Rep 62.

Interestingly, the Court of Appeal decided that disclosure of a previous expert's report should be construed to include any draft reports that were prepared before the final report was produced.

## IMPLICATIONS

In terms of good litigation practice, the judgment is important in these ways:

(a) litigators should appreciate that the permission of the court is not required to *instruct* an expert. Rather, the court has the power to restrict expert evidence, which means that a party may find he cannot *rely* on that expert evidence;
(b) an order of the court that does not refer to the name of the party's expert, only to his or her discipline, must be construed as giving that party the ability to instruct a separate expert, in accordance with this judgment;
(c) if, in other circumstances, it was appropriate for the court to give permission to a party to rely upon further expert evidence, then it is likely any such order will be conditional upon that party disclosing the first expert's report (including any drafts) to the other party first; and
(d) it is advantageous for a party to instruct the expert to prepare a report (or a draft giving the substance of the advice) first so that the party will know, by the time of the first case management conference, whether he will rely on that evidence. However, this is subject to a real risk that the court may not allow that party to rely on that expert report. In these circumstances, a client should always be advised of this risk and counselled appropriately about being unable to recover the cost of obtaining the report from the other party.

ANDREW ALONZI*

*BA (Hons), PGCHE, solicitor and Senior Lecturer, Nottingham Law School, Nottingham Trent University.

# LOSS OF A CHANCE IN CLINICAL NEGLIGENCE: POLICY CARRIES THE DAY

*Gregg v Scott* [2005] UKHL 2

*(Lord Nicholls of Birkenhead, Lord Hoffmann, Lord Hope of Craighead, Lord Phillips of Worth Matravers and Baroness Hale of Richmond)*

The question whether to impose a duty of care in negligence (particularly where the prospective defendant is a public authority) is a fertile area of litigation in which the influence of the European human rights jurisprudence is keenly felt.[1] The development of a stable approach to rules governing causation of damage is no less pressing and problematic.

The latest decision on the subject of the House of Lords in *Gregg v Scott* raises a number of legal difficulties that appear to have been resolved as a matter of practice by policy considerations.[2]

## BACKGROUND FACTS

The claimant consulted the defendant GP about a lump under his left arm on 22 November 1994 and was negligently told that it was benign. He moved home and consulted his new GP on 22 August 1995 and was promptly referred to hospital on suspicion that the lump was cancerous. He was seen at hospital on 2 November 1995 and cancer was confirmed by biopsy.

Whilst pathological tests were being carried out he was admitted to hospital as an emergency on 13 January 1996 as a result of the spread of the cancer into his left pectoral region. He had chemotherapy and then, since there was incomplete response, high-dose chemotherapy involving stem cell replacement and was discharged in September 1996. In early 1998 he developed a tumour under his right arm and was given palliative chemotherapy and told that he could not be cured. After another suspected relapse in April 1998 he was given further palliative chemotherapy. Since then there had been no recurrence of the disease.

## PROCEDURAL HISTORY

The claim was originally for deprivation of a full and early cure by the negligent delay in referral. Before trial, however, it became clear that the cancer was of a rare type with a worse prognosis so that the claimant claimed in the alternative for reduction of his chance of a favourable outcome. The trial judge found that without the nine month delay in diagnosis and treatment of the cancer the claimant would probably have achieved remission without high dose chemotherapy and that the delay had reduced his prospect of survival for ten years from 42% (when he first consulted the defendant) to 25% at the date of trial. Survival for ten years is used by the medical profession as a proxy for cure.

* Solicitor, Professor of Litigation, Nottingham Law School, Nottingham Trent University.

[1] See, for example, the effect of *Osman v United Kingdom* [1999] 1 FLR 193 and *Z and Others v United Kingdom* [2001] 2 FLR 612 on the principles laid down in *X v Bedfordshire County Council* [1995] 2 AC 633.

[2] [2005] UKHL 2, [2005] 2 WLR 268.

The judge found against the claimant on the basis that it was not probable that, absent the breach, he would have had a better outcome (whose chance without the negligence of the defendant was only 42%, *ie* less than evens). The Court of Appeal dismissed his appeal by a majority (Latham LJ dissenting).[3]

## LEGAL BACKGROUND

The law has traditionally required a past fact (actual or hypothetical) to be proved on the balance of probabilities. At the stage of quantification of loss arising out of the proved past facts, however, courts may and routinely do estimate the prospects of uncertain future events occurring on a percentage scale of probability. Where the defendant's wrongful conduct has deprived the claimant of a financial opportunity courts have in some cases been prepared to award damages proportionate to the likelihood of that opportunity being realised.[4] The importance of *Gregg* is that the claimant sought to import into a question of a past hypothetical outcome a proportionate approach to liability and quantification of loss by reference to statistical evidence, direct evidence of the putative outcome, absent the defendant's negligence, being *ex hypothesi* unavailable.

## CLAIMANT'S ARGUMENTS BEFORE THE HOUSE OF LORDS

Before the House of Lords (which dismissed his appeal by a three to two majority[5]) the claimant advanced two main arguments. He said that the growth of the tumour during the period of delay with the additional pain it caused was actionable damage and the loss of prospect of recovery was consequential upon that physical injury and therefore recoverable. Alternatively, he argued that he was entitled to compensation for the loss of the prospect of a cure, rather than for the loss of the cure itself (on the basis that he could not establish that a cure had been probable, absent the breach).

These arguments were suggested as an agenda for the appeal in the Appendix to Stapleton's celebrated article "Cause-in-Fact and the Scope of Liability for Consequences".[6] That article, in a rigorous and systematic analysis, proposes that factual questions relating to the involvement between tortious conduct and damage suffered should be kept separate from normative questions whether the defendant should be held responsible for the consequences of that tortious conduct.

## THE DECISION OF THE HOUSE OF LORDS

The decision of the House comprises five speeches giving different weights and answers to the two arguments and conflicting views on the dictates of policy. This, and the

[3] [2002] EWCA Civ 1471.

[4] *Chaplin v Hicks* [1911] 2 KB 786; *Kitchen v Royal Air Force Association* [1958] 1 WLR 563; *Allied Maples Group Ltd. v Simmons & Simmons* [1995] 1 WLR 1602.

[5] Lord Hoffmann, Lord Phillips of Worth Matravers and Baroness Hale of Richmond in the majority; Lords Nicholls of Birkenhead and Hope of Craighead in the minority.

[6] (2003) 119 LQR 388–425.

different interpretations of earlier authorities, render a single *ratio* hard to discover, and for that reason the factors and arguments swaying each member of the House require some detailed examination.

*The first argument*

Lord Nicholls dismissed this shortly as "superficially attractive ... but would not ... answer the fundamental issue" *ie*, whether in a clinical negligence case a chance of recovery less than 50% lost by negligence should be compensated.[7] This seems an inadequate response although it is possible to infer that the first argument assumed a lesser significance during the hearing.

Lord Hoffmann rejected the argument because, for the physical damage (the early spread of cancer) to found a claim for reduction in the prospects of survival, a reduction in the chance of survival must be a recoverable head of damages.[8] If, on the other hand, the claim was for depriving the claimant of his chances of survival, it must fail because the judge had found that the early spread of cancer did not on the balance of probabilities cause his failure to survive. The difficulty was that uncertain future outcomes could only be compensated on a proportionate basis where, if they eventuated, they would on the balance of probabilities have been attributable to the breach of duty on the part of the defendant. Whereas in the routine claim in respect of post-traumatic arthritis or extra earnings recoverable on a missed promotion, such as that in *Doyle v Wallace*, there was no dispute that the loss was attributable to the defendant's fault, in this case the judge concluded the likelihood of such attributability was less than 50%.[9]

Lord Hope appears in his conclusion to have treated the argument as a safety net.[10] For him the characterisation of the pleaded case as one of a claim for loss of a chance was inaccurate: it was rather, in his view, one of negligence causing a variety of forms of loss all flowing from the enlargement of the tumour: a physical injury.[11] The reader may regret the factual exigencies that led to such ambiguity in a case that was designed to establish an important principle.

Lord Phillips confined himself to pointing out the differences in approach between Lords Nicholls and Hope: the latter only would have upheld the first argument.[12]

Baroness Hale rejected the argument on the ground that the judge had not found that the delay in treatment caused the spread of the cancer. If it had been so caused, damages for consequential loss or reduction in the chance of a cure would have been available on a conventional assessment of that chance. In this she agreed with Lord Hoffmann's view that, to obtain a proportionate recovery of a potential future loss, the claimant must prove the damage that gave rise to the loss on the balance of probabilities.[13]

Since the judge had found that the claimant would probably have suffered the same problems at different times, absent the negligent delay, all he could have claimed (had he chosen that approach) was any extra pain, suffering and loss of amenity resulting from the *delay* in treatment (rather than the deterioration of his condition).

[7] *Op cit*, note 2 para 58 referring back to paras 1-4. All subsequent paragraph numbers refer to this citation.

[8] Para 71. There was no finding to that effect by the Court of Appeal.

[9] [1998] PIQR Q146.

[10] "If it is necessary to prove that this loss was caused by a physical injury, the enlargement of the tumour which the negligence caused was such an injury." Para 121.

[11] Para 117.

[12] Para 175.

[13] Para 200, *cf* text at note 9.

*The second argument*

Lord Nicholls would accept the argument since the uncertainty of medical outcomes meant that a loss of prospects was an appropriate characterisation of damage, as the doctor's duty was to promote the prospects of recovery.[14] Justice required that the remedy should be provided just as it was in claims for financial loss.[15] The law could resort to statistics in order to leap evidentiary gaps.[16]

Lord Hoffmann concluded that the House had in *Hotson* and *Wilsher* decided against allowing recovery on the basis of likelihood of cause and effect being less than 51%.[17] In *Fairchild*[18] the House had constructed a special rule for application in restrictively defined circumstances imposing liability on a defendant whose act or omission caused a material increase in the risk of the adverse outcome which it was his duty to prevent.

The spread of the claimant's cancer was not random but determinate, even if medical science was unable to illuminate the cause. Thus it should not be treated as an emanation of pure chance.[19] The law might properly compensate on a proportionate basis the loss of a chance of a future outcome that depended on the intervention of a third party, since the determinate cause was influenced by that intervention.[20]

In essence, the claimant was asking the House to adopt the claimant's position in *Wilsher* although *Wilsher* had itself been approved in *Fairchild.* To accede to the claimant's argument would require both *Hotson* and *Wilsher* to be overruled and the limitations of the rule in *Fairchild* to be removed.[21]

Lord Hope chose to see the loss of a significant prospect of a successful outcome as "one element among several in the claim for which there is a single cause – the enlargement of the tumour."[22] In that sense he appears to have preferred the claimant's first argument.

Crucially for Lord Phillips, the claimant was alive nine years after he should have begun treatment and nearly eight years after he in fact did begin. In that respect he had the advantage over the trial judge of four years hindsight of the claimant's actual medical history. Thus, he was able to conclude that the delay in beginning treatment probably did not affect his prospects of surviving, although they probably did cause the spread of the cancer and the relapse with additional pain and apprehension. The difficulties encountered in evaluating loss of a chance rather than employing a balance of probabilities test were a policy factor against the introduction of a right to compensation for the loss of a chance.

Baroness Hale also concluded that the outcome of the argument depended on policy questions.[23] These are dealt with below.

*Relationship with Hotson*

In *Hotson* the claimant's injury was determinate at the time of the doctor's negligent omission to treat him. The House was able, therefore, to approach his case on the

[14] Para 24.

[15] *Allied Maples*, see note 4.

[16] Para 31.

[17] *Hotson v East Berkshire Health Authority* [1987] AC 750; *Wilsher v Essex Area Health Authority* [1988] AC 1074.

[18] *Fairchild v Glenhaven Funeral Services Ltd* [2002] UKHL 22, [2003] 1 AC 32.

[19] Paras 79–80. "What we lack is knowledge and the law deals with lack of knowledge by the concept of the burden of proof."

[20] Para 82.

[21] Para 85.

[22] Para 118.

[23] Para 212.

conventional basis of balance of probabilities in relation to past actual facts.[24] According to all but Lord Hoffmann and Lord Phillips the House in *Hotson* left open the position in the case, such as the instant case, where the cause of the injury was unresolved.[25] Both ignore the distinction and regard the uncertainty of outcome as fatal to the claimant's task of satisfying a balance of probabilities standard of proof, [26] although Lord Phillips towards the end of his speech appears to consider that a claim for increase in the chance of an adverse outcome might be more properly made on a proportionate basis *after* the adverse outcome had eventuated.[27] This must be read as limited to the case where the claimant has proved causation by the defendant of the underlying injury on the balance of probabilities.

*Relationship with Fairchild*[28]

This case was also prayed in aid by both the majority and the minority. For Lord Nicholls, it was evidence of the willingness of the courts to "adapt their processes so as to leap an evidential gap when overall fairness plainly so requires".[29] According to Lord Hoffmann, *Fairchild* provided a special rule, restrictively defined, imposing liability for conduct which increased the chances of the claimant contracting mesothelioma.[30] It is not clear, however, why conduct increasing a chance of disease should be treated differently from conduct depriving a claimant of a chance of recovery from an existing disease.

Lord Phillips, however, showed little enthusiasm for the "change in the law in the interests of justice" introduced by *Fairchild* and warned of the danger, if special tests of causation are developed piecemeal, of the destruction of the coherence of the common law.[31] One wonders whether this sentiment extends to disapproval of *Fairchild* itself. If the point is that special cases should be limited, criteria for their imposition and denial would have been welcomé.

The most explicit reason for refusing to extend *Fairchild* came from Baroness Hale who held that *Fairchild* and *Chester*[32] were remedies for particular problems that did not alter the principles applicable to the great majority of personal injury claims.[33] This is a cogent point but the precise limits on *Fairchild* are undefined and practitioners seek in vain for the certainty and coherence that Lord Phillips sought to protect.[34]

*Relationship with Chester*

Baroness Hale was the only member of the House to refer to *Chester*. This was in the context of the need to show actionable damage: in *Chester* the damage was clear.[35] If that is to be contrasted with the doubts over the prognosis for the claimant in *Gregg*, the point goes back to the resolved/unresolved dichotomy. If, on the other hand, it is

[24] In a lecture at the British Institute of International and Comparative Law on 9 May 2005, Professor Jane Stapleton proposed describing such cases as "resolved" and those where the future for the claimant was uncertain at that time as "unresolved".

[25] The majority on this point at paras 38, 108 and 212; the minority at paras 75 and 174.

[26] Paras 75 and 174.

[27] Para 190.

[28] *Fairchild v Glenhaven Funeral Services Ltd* [2003] 1 AC 32.

[29] Para 31.

[30] Para 78.

[31] Para 172.

[32] *Chester v Afshar* [2004] 3 WLR 927.

[33] Para 192.

[34] For a penetrating analysis of *Fairchild* and the problems unresolved by it, see J Stapleton, "Lords a'leaping Evidentiary Gaps" [2002] *Torts Law Journal*, 277–304.

[35] Para 217.

a point going to rules for the proof of causation, it is unclear how the clarity of the damage is a crucial factor. To be sure, there was no doubt what damage Ms Chester suffered; the question was whether that damage had been caused by the breach of the defendant.

The claimant in *Gregg* could be forgiven for wishing that his case had been heard with or by the same tribunal as that in *Chester* and for wondering why the concentration on the protection of patient autonomy and dignity in the latter case did not receive a mention in his own case.[36]

*Policy questions*

While Baroness Hale and Lord Phillips explicitly acknowledged that the question whether to provide a remedy for loss of a chance was a policy question, the other members of the Judicial Committee, except Lord Hope, clearly used policy considerations in arriving at their conclusions.

Lord Nicholls appealed to the need for the common law to develop so as to provide fairness and not arbitrariness.[37] The "floodgates" concern of the majority of the Court of Appeal should not stand in the way of such development,[38] the more so since the principle contended for would not impinge on *Fairchild*.[39] Thus, a patient should have a remedy when he loses the very thing that the doctor's duty should have protected, and this should apply as much to indifferent as well as to probable prospects of recovery.[40] The first part of this argument did indeed appear to attract the House in *Chester*.

Oddly, Lord Nicholls sought to overcome the potential disadvantage of extra expense for the National Health Service (NHS) by confining a remedy to cases of "significant" diminution of a "reasonable" prospect of recovery on a proportionate basis whilst urging courts to avoid the spurious precision of percentage awards.[41] The difficulty of definition inherent in the need to confine the ambit of his preferred approach to cases with a "significant degree of medical uncertainty" was brushed aside as the type of determination courts are used to making.[42]

For Lord Hoffmann to impose a control mechanism limiting the recovery for loss of a chance to cases of injury was unprincipled and would lead to litigation over the definition of "injury".[43] Nor would confining recovery to cases where inability to prove causation resulted from lack of medical knowledge (as in *Wilsher*) rather than lack of knowledge of the facts (as in *Hotson*) be principled, particularly since in *Hotson* the claimant's inability to prove whether his hip could have saved resulted from the negligence of the defendant in failing to examine him.[44]

Finally, Lord Hoffmann appealed for certainty in the law. A wholesale substitution of possibility for probability as the causal standard would be tantamount to legislation with enormous consequences for insurers and the NHS.

Lord Phillips set himself the task of examining the practical consequences of the proposed change in the law and in so doing set about a detailed review of the statistical evidence in the case. This is an unusual approach in the House and Lord Phillips

[36] See, for example, Lord Steyn at para 18, Lord Hope at para 77 and Lord Walker at para 92.
[37] Para 45.
[38] Para 48.
[39] Para 51.
[40] Para 42.
[41] Para 54.
[42] Para 50.
[43] Para 87.
[44] Para 88.

subjected the evidence of the claimant's expert evidence to severe scrutiny (the detail of which is beyond the scope of this comment) and cast doubt on the use of statistical evidence in general and its use and interpretation in this case in particular. The detail of the criticism particular to the case is hard to follow without the evidence itself or the transcripts of what was said about it. The result of the analysis was to show the difficulty of accurate assessment of the loss of a chance in clinical negligence cases. That difficulty, compared to the simplicity of the probability standard, was a policy factor weighing against substituting the former for the latter.[45]

For Baroness Hale to accept loss of a chance as the damage which founded a right of action (whether or not accompanied by a physical change in the claimant's condition) would lead to the prospect of liability in almost every case. Tort law should not, however, be used for punishment of carelessness but for compensation of wrongs probably caused.[46] It was not irrational to treat medical practitioners more favourably than other practitioners: one may expect to have money or the chance of having money interchangeably. There is, however, a real difference between a disease-free state and the chance of having that state.[47]

Almost any claim for a lost outcome could be reformulated as a claim for the loss of a chance of that outcome.[48] If the law moved in that direction a claimant with a 55% chance of a cure would receive 55% rather than 100% of her loss whilst a claimant who had lost a 20% chance would recover 20%. Compared with outcomes under a balance of probabilities test winners would suffer and losers would gain. And, if a claimant were allowed to treat loss of outcome and loss of a chance of outcome as alternatives, a claimant with a strong case would opt for the former (and recover 100%) and one with a weak claim for the latter (and recover say 20%). This would clearly be unfair on defendants.[49] The decision in *Allied Maples* received little attention from their Lordships despite confusion at large about its precise *ratio* and the sentiment that the law should value the person more highly than property.[50]

In most personal injury claims (which are straightforward) claimants would suffer by the introduction of proportionate liability and recovery would become unpredictable with more complex negotiations and trials. The costs of such a change would outweigh the benefits.[51] The approach to causation should be the same for both past and future events.

*Quantum*

It is unusual for eight months to elapse between argument before the House of Lords and the delivery of the judgments in the case. That such a period of time elapsed in this case is a mark of the difficulty of the questions at issue. Some months after argument had been concluded the parties were requested to make written submissions on how damages should, if the claim succeeded, be assessed. This must in itself have raised the claimant's hopes and the opening of Lord Hope's speech suggests that he expected, when beginning to draft it, to be in the majority.[52]

[45] Para 170. "A robust test which produces rough justice may be preferable to a test that on occasion will be difficult, if not impossible to apply with confidence in practice."

[46] Para 217.

[47] Para 220.

[48] Para 224.

[49] Para 225.

[50] See notes 4 and 15.

[51] Para 224.

[52] Para 92.

Whilst the question of quantum did not, in the event, arise, by inference the proper approach to it was not agreed between their Lordships. Lord Phillips noted with a hint of satisfaction that Lord Hope would award full damages for pain, suffering and loss of amenity whereas Lord Nicholls appeared to favour recovery proportionate to the chance lost.[53]

## CONCLUSION

This comment does not do justice to the detail of the judgments but may give a flavour of the diversity of the reasoning employed by their Lordships. The outcome is that a claimant in a clinical negligence or personal injury case cannot recover damages for less than a 51% chance of obtaining a more favourable outcome than he would have had in the absence of negligence on the part of the defendant.

A substantial part of the judgments themselves are taken up with a close analysis of the evidence in the case. That is a far from ideal basis on which to decide an important doctrinal issue, the more so since that evidential complexity is one policy ground leading the majority to retain the requirement for a balance of probability standard. The reasoning of Baroness Hale, nevertheless, is persuasive in relation to the volume of "small chance" proportionate claims that the establishment of a right to sue for a lost chance less than 50% (assuming negligence had been established) would have unleashed.

That is not, however, to ignore the force of Lord Nicholls "hollow duty" argument. Healthcare providers acting rationally could now decide against the allocation of scarce resources to patients whose chance of a better outcome, after receiving such resources, did not reach 50%. Evidence-based medicine would make simple the composition of lists of common indications for medical intervention where the prospect of a favourable outcome exceeded and fell short of 50%. Although, in the real world, ethical and moral considerations would prevent such an approach, it may be thought undesirable that in the light of the outcome of *Gregg* it would not attract condemnation by the law of tort.[54]

The justification of the counter-intuitive disparity of treatment between claims involving economic loss and those involving injury to the person may not satisfy those who seek to vindicate the right to receive non-negligent medical treatment. To the extent that the instant case and *Chester* were seen to be related, they may regret that the approach of the majority in the latter case did not prevail in the former.

MARK MILDRED*

[53] Para 175.

[54] A point brought out by Professor Richard Butler in his helpful comments on a draft of this note.

# THE STATUS OF THE MODERN APPRENTICESHIP

*Flett v Matheson* [2005] IRLR 412

*Employment Appeal Tribunal*

*(Mr Justice Burton, Sir William Morris KBE OJ, Mr R N Straker)*

## INTRODUCTION

From the 1960s onwards, the uptake of traditional apprenticeships was in decline. The introduction of the Modern Apprenticeship in 1993 was an attempt to revive this tradition and extend the approach to emerging sectors and a wider range of occupations.[1] Some of the more successful features of the traditional apprenticeship were to be maintained, including the stipulation that there should be a written agreement or pledge between the employer and apprentice, specifying rights and obligations.[2]

In contrast to the traditional apprenticeship however, the Modern Apprenticeship was to involve not only the employer and the apprentice, but also the Learning and Skills Council (the "LSC"), which would provide supervision for the apprentice's training, and (usually) a third party training provider.[3] The apprentice would work for the employer and receive training leading to a qualification such as a National Vocational Qualification (NVQ). Within the Modern Apprenticeship framework, the young person would be paid a wage, at a level determined by the employer and apprentice.[4]

In terms of statute law, a contract of apprenticeship has been raised to the status of a contract of employment.[5] However the traditional (or medieval) concept of a contract of apprenticeship remains a distinct entity at common law. The essential feature of an apprenticeship is that the apprentice contracts to be taught a trade or calling; the execution of work for the employer is only a secondary consideration. Furthermore, the contract is for a fixed term and not terminable on notice, for example in the event of redundancy due to a downturn in work.[6] The level of compensation may also be different for termination of a contract of apprenticeship as opposed to a contract of service. Under a contract of apprenticeship, an apprentice who is wrongfully dismissed may then have a claim for enhanced damages by reason of loss of his or her prospects as a tradesperson.[7]

In *Flett v Matheson*,[8] the Employment Appeal Tribunal ("EAT") decided that the Modern Apprenticeship arrangement was a contract of employment as opposed to a traditional contract of apprenticeship. The decision of the EAT is interesting not only

[1] Learning and Skills Council (2004), "21st Century Apprenticeships: End to End Delivery of the Modern Apprenticeships", Department for Education and Skills, p9.

[2] H Gospel, "The Revival of Apprenticeship Training in Britain" (1997) *Centre for Economic Performance Discussion Paper* 372, p10.

[3] PLC Employment (2005), "Modern Apprenticeships: employment status", Practical Law for Companies: http://www.practicallaw.com/3-200-6334.

[4] H Gospel, "The Revival of Apprenticeship Training in Britain" (1997) *Centre for Economic Performance Discussion Paper* 372, p10.

[5] See, for example, s 230 (1), Employment Rights Act 1996.

[6] In *Wallace v CA Roofing Services Ltd* [1996] IRLR 435 the EAT held that the employer could not terminate the contract of apprenticeship solely on the ground of redundancy falling short of closure or of a fundamental change in the character of the defendant's enterprise. Damages were awarded to the apprentice in this case.

[7] *Dunk v George Waller & Sons Ltd* [1970] 2 QB 163.

[8] [2005] IRLR 412.

in determining the status of the Modern Apprenticeship, but also in indicating the factors by which a contract of apprenticeship and contract of employment are identified in more general terms, and the rights of those holding such contracts.

## THE FACTS

Mr Flett worked for EAB Electricals ("EAB") for eight months in 2002 before entering into a Modern Apprenticeship. The Modern Apprenticeship was drawn up in the form of an individual learning plan ("ILP"), and consisted of a tripartite agreement between Mr Flett, EAB and a training provider called JTL (a government-funded and supported body under the ægis of the LSC and the Joint Industry Board (the "JIB")) responsible for providing training to apprentices in the electrical field. The ILP referred to a "learning start date" of 10 September 2002 and stated that the "duration of work" would be 42 months. The relationship between the parties was also underwritten by the Joint Electricity Board Training Scheme for Electrical Installation Apprentice Craftsman and Technicians 1999 (the "Training Scheme"). This set out the specific industry standards to be met by the apprentice for qualification in his field of work.

EAB's business was taken over by Mr Matheson, and Mr Flett's contract was terminated. Mr Flett claimed unfair dismissal and breach of contract against Mr Matheson, arguing that EAB's liabilities transferred to Mr Matheson on transfer of the business under the provisions of the Transfer of Undertakings (Protection of Employment) Regulations 1981. He argued that he was employed under a contract of apprenticeship of a fixed term, which would allow him to recover not only damages for the failure of the employer to complete the contract to the end of the four year term, but would also make him eligible to pay damages for loss of benefit of the training during the apprenticeship and continuing loss of the opportunity for Mr Flett to go out into the market qualified in his trade.[9] Mr Flett accepted that if his contract was found to be one of employment rather than apprenticeship, he would only be entitled to one week's notice pay.

At first instance, the Tribunal found that there was no contract of apprenticeship, and also no contract of employment. As a result, there was no jurisdiction to entertain Mr Flett's claim for breach of contract, as there was no contract falling within the definition of the Employment Tribunals Act 1996, section 42(1) (although the Tribunal concluded that there was a contract of some sort between EAB and Mr Flett). That section is the interpretation section of the Employment Tribunals Act 1996, and provides that a "contract of employment means a contract of service or apprenticeship, whether express or implied, and (if it is express) whether oral or in writing".

Accordingly the case was dismissed by the employment tribunal at first instance. Mr Flett appealed to the EAT.

[9] Mr Flett's argument was based on that in *Dunk v George Waller & Sons Ltd* (*op cit*). In that case, the High Court held that the apprentice could claim not only loss of earnings, but also damages for loss of the skills he would have attained as a result of the training, and loss of status on the labour market through the lack of a qualification.

## THE JUDGMENT AND ITS EVALUATION

*Identifying a traditional apprenticeship*

The first issue that the EAT had to consider was whether or not the Modern Apprenticeship was analogous to the traditional contract of apprenticeship. Although not specifically referred to in the judgment, it is useful to outline the features of a traditional contract of apprenticeship here. These features are summarised in the judgment in *Thorpe v Dul (No 1)*,[10] where four elements of the traditional contract of apprenticeship were identified:

(a) the employer receives the apprentice into his service for a fixed term;
(b) the apprentice acquires, under the employer's control, a practical knowledge of the employer's trade or business;
(c) the employer pays the apprentice; and
(d) the apprentice agrees to be subject to the employer's conditions of service.

In the present case, Burton J referred to the case *Wallace v C A Roofing Services Limited*[11] to identify the elements of the traditional contract of apprenticeship (as distinct from the Modern Apprenticeship). In *Wallace* the claimant was employed as an apprentice sheet metal worker prior to the introduction of the Modern Apprenticeship scheme. The apprentice was employed on a fixed term contract, a classic feature of the traditional contract of apprenticeship. The oral agreement between the parties did not include a term that the period of training might be terminated prematurely (on the ground of redundancy, for example). There was no evidence in *Wallace* of any tripartite agreement. The training of the apprentice was therefore largely under the employer's control, save for day release one day a week to study for his City and Guilds exams. There were further features of the traditional contract of apprenticeship: the apprentice was paid by the employer; and written terms and conditions were in place between the employer and apprentice.

*Distinguishing the Modern Apprenticeship from a traditional apprenticeship*

There have been instances in case law in which the Modern Apprenticeship has been considered analogous to the traditional contract of employment. In the case of *Whitely v Marton Electrical Limited*,[12] it was held that the tripartite agreement between the employer, apprentice and the local Training and Enterprise Council ("TEC") was a contract of apprenticeship. This decision was based on the conclusion by the EAT that the employer had wider responsibilities than he would have towards an ordinary employee, and that the contract was for the "duration of the training plan" (*ie*, a fixed term). Burton J stated that the EAT disagreed with this particular finding, on the basis that the employer did not have sufficient control over the education and training of the apprentice, a feature central to the traditional contract of apprenticeship (see below).

In his judgment, Burton J made it clear that he did not intend to come to a conclusion about the status of every Modern Apprenticeship, but identified three elements of the traditional contract of apprenticeship which were absent from the present Modern Apprenticeship scheme:

[10] [2003] ICR 1556.
[11] [1996] IRLR 435.
[12] [2003] IRLR 197.

(1) The apprenticeship was for a fixed term.
(2) As a result of the apprenticeship being for a fixed, usually four-year term:
(a) The employer agreed to employ the apprentice for such fixed term, or for the duration of the apprenticeship;
(b) The employee agreed to remain apprenticed for such period;
in both cases subject to certain, limited, get-out arrangements.
(3) Fundamentally, the employer had an obligation to educate and to train, and effectively to secure, subject to the get-out arrangements on both sides, the required qualification for the apprentice to take into his full time career.

In the ILP, in the *Flett* case, although it was stated to be for a duration of 42 months, that was not a fixed term. The completion date was only "anticipated" as the apprentice could only complete when the Joint Industry Board issued him with a completion certificate and an ECS Card/Joint Industry Board Grade Card.[13] Further, the employer had the ability to terminate the apprenticeship during that term if the apprentice was unable to attain the required industry standard to progress through stages of the Training Scheme.[14] Most importantly, it was not the employer's responsibility to provide sufficient training for the employee to become qualified. What the employer was bound to do was provide a full opportunity for access to education and training to the apprentice, to allow assessment of the trainee while at work, and usually, pay the apprentice whilst he or she attended a course or college. The EAT therefore decided that the Modern Apprenticeship did not constitute a traditional contract of apprenticeship in this instance.

*Rights under traditional apprenticeship contracts*

Apprentices have broadly the same rights under statute as those persons under a contract of employment. For example, they have the protection of anti-discrimination laws.[15] However, there is a significant advantage to claimants in proving that they hold a contract of apprenticeship rather than a contract of employment and that is the qualification for rights under common law. A contract of apprenticeship is not terminable at will, but is, as we have seen, for a fixed term. It is terminable only in exceptional circumstances. This has a massive impact, for example, on the ability of employers to make apprentices redundant: it is not possible for apprentices to be made lawfully redundant unless the employer's business has changed so fundamentally that the apprentice could no longer be taught the trade for which he or she was engaged.[16]

Similarly, only serious misconduct will justify dismissal. In *Thorpe v Dul (No 1)*,[17] the tribunal employed the common law summary dismissal gross misconduct test in deciding whether the dismissal of an apprentice was justified.

Further, there is potentially a huge difference in the remedies available for the breach of a contract of apprenticeship as opposed to breach of a contract of employment or a contract for training. In the judgment in *Dunk v George Waller & Sons Limited,*[18] Widgery J decided that:

[13] Clause 3.9 of the Training Scheme.

[14] Clause 3.13, *ibid.*

[15] See, for example, s 4 (2), Race Relations Act 1976 ("RRA"), and s 6 (2), Sex Discrimination Act 1975 ("SDA"). On the other hand, there are circumstances in which they may not be entitled to the minimum wage (Regulation 12(2) National Minimum Wage Regulations, SI 1999/584).

[16] *Wallace v C A Roofing Services Ltd* [1996] IRLR 435.

[17] *Op cit* n 10.

[18] [1970] 2 QB 163.

> a contract of apprenticeship is significantly different from an ordinary contract of service if one is to consider damages for breach of the contract by an employer. A contract of apprenticeship secures three things for the apprentice: it secures him, first, a money payment during the period of apprenticeship, secondly, that he shall be instructed and trained and thus acquire skills which would be of value to him and thirdly it gives him status in the labour market.

The Court of Appeal therefore decided to compensate the apprentice for his loss during the remaining period of his four year apprenticeship (the short-term loss) and his loss after expiry of this term (the long-term loss). Compensation for long-term loss consisted of a sum representing the difference in earnings *per* week between a qualified and non-qualified apprentice: recompense for his loss of training and status on premature termination of his contract.

In *Flett*, the claimant put forward a claim for £50,459·83 on the basis of a breach of a contract of apprenticeship by EAB. In contrast, if the apprentice were only employed under a conventional contract of employment, he would only be entitled to one week's notice, a sum of £112·12, representing a huge difference in liability for the employer.

*Contract for training*

It has been noted above that the primary purpose of the contract of apprenticeship is training, and that the work done for the employer is only secondary. The apprenticeship, as an ancient form of on-the-job-training in the skilled trades[19] is only one type of contract in which the provision of training is central. The rights of persons under general contracts for training will, therefore, be examined to compare and contrast the position of the apprenticeship.

In *Wiltshire Police Authority v Wynn*,[20] the Court of Appeal had to decide the employment status of police cadets. It found that the primary purpose of the cadetship was to teach and learn, but distinguished the cadetship from a contract of apprenticeship. In particular, Waller J noted that, unlike the apprentice under a contract of apprenticeship, the police cadet did not gain any qualifications (*ie*, as a police constable) on completion of the cadetship, and for that reason the cadet could not hold a contract of apprenticeship. The Court of Appeal also found that the police cadet was not employed under a contract of service, as the duties carried out by the cadet were not sufficiently extensive and constituted only "minor assistance" to the police force. Accordingly, the police cadetship was in a class of its own.

The position of a pupil barrister was considered in the case of *Edmonds v Lawson*.[21] The Court of Appeal had to decide whether the pupil was employed as a "worker" under the provisions of the National Minimum Wage Act 1998 ("NMW"). Under section 54(3) of the NMW a "worker" is a person who works under

> (a) a contract of employment [which includes a contract of apprenticeship] or (b) any other contract, whether express or implied and (if it is express) whether oral or in writing, whereby the individual undertakes to do or perform personally any work or services for another party to the contract whose status is not by virtue of the contract that of a client or customer of any profession or business undertaking carried on by the individual.

The Court of Appeal held that the pupil barrister was not employed under a contract of apprenticeship. In coming to this conclusion, the Court of Appeal referred to the

[19] A Hole, "Learn as you earn" (2002) 152 (7027) *New Law Journal* 558.

[20] [1980] ICR 649.

[21] [2000] QB 501; 2000 WL 191185.

fact that the pupil barrister was not under the control of her master as an apprentice would be, and in particular was not under any duty or obligation to do anything not conducive to her own training and development. She was unpaid by the chambers, reflecting the lack of expectation that she would render any services of value to the chambers; although there was a provision in the Bar Code of Conduct that as between the pupil and master she would be paid for any work of value that she did in fact carry out.

In *Edmonds v Lawson*, a binding contract was identified, which was neither an orthodox contract of employment nor a contract of apprenticeship, nor was it "any other contract, whether express or implied, whereby the individual undertakes to do or perform personally any work or services for another party to the contract". It was a contract for the provision of education and training, a finding which was potentially very significant for both parties. The chambers took on the pupil with a view to awarding the pupil a tenancy at the end of the training period, and the pupil gained the work experience necessary to allow her to practise in her chosen field.

*Distinguishing the Modern Apprenticeship*

In looking at the Training Scheme, it is clear in the *Flett* case that the employer did have to make a contribution to the training of the apprentice, and, crucially, he was able to terminate the employment of any apprentice who was unable to attain the required industry standard (section 3·13 of the Training Scheme). The employer was obliged to ensure that the JIB Apprentice Logbooks were correctly and accurately completed in accordance with the requirements detailed in the Logbook and could only apply for termination under section 3·13 of the Training Scheme if he had "met the requirements of full involvement in and commitment to the apprentice's training and assessment process and after due consultation with, and warnings to the apprentice".

In *Flett*, Burton J referred to the Modern Apprenticeship as a whole as a "training contract", but made it clear that the *training obligation* lay squarely with the training provider and not with the employer. Although no finding was made by the EAT as to the existence or otherwise of a contract for training between the employer and apprentice, Burton J stated that the employer was not obliged to see that the apprentice qualified in his or her trade, and therefore had no training obligation towards the apprentice. All that the employer had to do was ensure that the apprentice had sufficient access to training for the duration of the contract and allow assessment of the apprentice whilst at work.

However, in neither *Edmonds v Lawson* nor *Wiltshire Police Authority v Wynn* was it the responsibility of the trainer to ensure qualification of the trainee. Indeed, contracts for training can be based solely on the provision of work experience and a "general education"[22] and so it is quite possible that the relationship between the employer and apprentice in the present case could form the basis of a contract for training.

*Rights under a contract for training*

Even if the relationship between the employer and apprentice in the present case could have formed the basis of a "contract for training", this is largely irrelevant if there is a contract of employment between the parties. A contract of employment carries with it far more rights and obligations, and would supersede a contract for training. The

[22] *Per* Denning MR in *Wiltshire Police Authority v Wynn* [1980] ICR 649.

existence of a contract for training would therefore only be relevant where there is no contract of employment: for example, in an analysis of the relationship between the training provider and apprentice (with which we are not concerned in this case).

In terms of rights, a person holding a contract for training will not come within the definition of "employee" under the Employment Rights Act 1996 ("ERA"), section 230 and therefore will not have the right to claim unfair dismissal or redundancy payments under the ERA. Under the legislation originally enacted, persons holding contracts for training did not have the benefit of the protection of anti-discrimination legislation. This is demonstrated by the case of *Daley v Allied Suppliers Limited.*[23] The claimant in this case was taking part in a work experience scheme as part of the Youth Opportunities Programme. She claimed race discrimination against the company for which she worked, but failed as she was deemed not to be employed within the meaning of the Race Relations Act 1976 ("RRA"), section 78. In response to this case, the RRA and the Sex Discrimination Act 1975 were amended so as to include sponsors of work experience schemes within the list of providers of vocational training.[24] However, in many cases, persons working under a contract for training will come within the definition of "worker" under the ERA enabling them to claim, for example, unlawful deduction from wages (section 13 ERA). The Health and Safety at Work Act 1974 covers persons undergoing "relevant training" (*ie*, work experience pursuant to a training course) who would otherwise not be covered due to their lack of employment status.[25] The Working Time Regulations 1998[26] and public interest disclosure provisions (sections 43 A-L ERA) are similarly extended to non-employed trainees.

*Contracts of Employment*

In the present case, the EAT held that the apprentice was employed under a contract of employment. Burton J remarked that the decision had been reached on the following basis:

> Every incident indicating employment seems, notwithstanding the paucity of evidence to have been present and to have continued during the training period; and the relevant documents redound with employer, employee and terms and conditions of employment. There is little doubt that the Appellant was working as an employee, and receiving wages as such.[27]

Unfortunately, the documents referred to by Burton J are not the subject of extensive analysis, but the Training Scheme, which is referred to, sets out a number of mutual obligations between employer and apprentice. For example, the employer is obliged to ensure that the JIB Apprentice Logbooks are correctly and accurately completed in accordance with the requirements detailed in the Logbook, as we have seen above. The payment of the apprentice also seems central to the finding of a contract of employment. Indeed, Burton J asserted that a possible reason that the tribunal at first instance in this case made the incorrect finding that there was neither a contract of employment nor a traditional contract of apprenticeship was that there had been a misunderstanding as to the remuneration of apprentices under the Modern Apprenticeship scheme. The tribunal followed the judgment of the EAT in *Thorpe v Dul (No 1)*[28] which found that there was no contract of employment between employer and

[23] [1983] IRLR 14.

[24] S 13 RRA and s 14 SDA.

[25] Ss 2–4, Health and Safety (Training for Employment) Regulations 1990, SI 1990/1380.

[26] SI 1998/1833, see s 42 ERA.

[27] *Op cit* n 8 at para 31.

[28] *Op cit* n 10.

apprentice as the employer was not responsible for the payment of wages. This was found to be a misunderstanding and when the case was resubmitted to the EAT, the EAT found in *Thorpe v Dul (No 2)*[29] that not only did the employer pay the apprentice but there was a contract of employment between the two parties.

*Rights of Employees*

As we have seen, some of the core legal protections only apply to employees, particularly the rights on termination of the contract of employment under the ERA. It is only employees who are entitled to receive statutory sick pay[30] and have rights to maternity and paternity leave.[31] Apart from the statutory rights, employers and employees have obligations that are implied into the contract between them, including the mutual duty of trust and confidence and the duty to act in good faith (imposed on the employee). The tax and social security treatment of a person providing services also depends on his or her employment status.[32]

In the present case, the terms and conditions of the original contract of employment between EAB and the apprentice determined that employees working for EAB were only entitled to one week's notice, until they worked for long enough to gain greater security. As a result of the way the present case was argued on behalf of the apprentice, the EAT's finding that the apprentice was employed therefore entitled him to compensation amounting only to £112·12.

However, the EAT thought that in an appropriate case it could be argued that the onset of a Modern Apprenticeship could override the provisions as to duration and terminability in the original contract of employment. The EAT considered the current provisions as to termination in the Modern Apprenticeship. According to the Training Scheme, the apprenticeship could be terminated in the event of the apprentice failing to make satisfactory progress with the training. If that were the case, the employer would have to apply to the training agent for transfer to another employer to continue the training. The EAT concluded that if the Modern Apprenticeship were to override the provisions of a contract of employment, then there would be an implied term in the Apprenticeship that the contract of employment was terminable on reasonable notice; such reasonable notice calculated with reference to the time within which it would be reasonable to expect the training agent in question to place the apprentice with another employer.

*Conclusion*

Following the case of *Flett*, it is likely that a Modern Apprenticeship consisting of a tripartite agreement under which the apprentice is paid by the employer but receives training from a training provider, will be deemed a contract of employment.

However, there are a myriad of possible arrangements which exist under the government-backed apprenticeships system[33] and the determination of the relationship between the parties within such a system will continue to depend on the facts of the case. If, for example, there is a definite fixed term under a Modern Apprenticeship scheme that is not subject to early termination by either party, and the employer still

[29] EAT/0041/04 & 0042/04.

[30] S 163(1), Social Security Contributions and Benefits Act 1992.

[31] Ss 4 and 13, the Maternity and Parental Leave Regulations 1999.

[32] For a full analysis of the tax and social security treatment of employed and non-employed persons see D Evans, (2005) "Employment Status", Practical Law for Companies: http://www.practicallaw.com/6-200-4244.

[33] PLC Employment, (2005) "Modern Apprenticeships: employment status", Practical Law for Companies: http://www.practicallaw.com/3-200-6334.

has some responsibility for training, then there is the possibility that this arrangement will be deemed a contract of apprenticeship. There is also certainly scope for determining that a contract for training exists between the employer and apprentice and/or the apprentice and training provider, depending on the training obligations between the parties.

As has been illustrated above, there is a significant difference between the rights and obligations of parties under contracts of employment, contracts of apprenticeship and contracts for training, and so the determination of the nature of the relationship between the parties under a Modern Apprenticeship scheme is very important. It is also likely that the continued success of the Modern Apprenticeship will depend on the determination of the nature these relationships. If, for example, a young person is deemed to have employed status during an apprenticeship, this will signal employer commitment and will make the Modern Apprenticeship more attractive to potential recruits in the future.[34]

LISA RODGERS*

[34] H Gospel, "The Revival of Apprenticeship Training in Britain" (1997) *Centre for Economic Performance Discussion Paper* 372, p10.

*BA (Hons), trainee solicitor, Edwards Geldard Solicitors.

NOTTINGHAM LAW JOURNAL

VOL 14(2) 2005

# BOOK REVIEWS

*Book reviews and books for reviewing should be sent to the address given at the beginning of this issue*

## "SPORTS LAW" OR "SPORT AND THE LAW"? A NECESSARY DEBATE?

*Sport: Law and Practice* by ADAM LEWIS and JONATHAN TAYLOR (eds), London, Butterworths, 2003, lxiv + 1162 pp, Hardback, £130, ISBN 0-406-94592-6

*Sports law and policy in the European Union* by RICHARD PARRISH, Manchester, Manchester University Press, 2003, 271 pp, Paperback £14·99, ISBN 0-7190-6607-7

In the preface to their work Lewis and Taylor assert that "[t]his book does *not* seek to address whether there is such a thing as sports law, still less whether there should be, because that is an issue of academic rather than practical interest."[1] To suggest that such a question is of only academic interest seems curious, at least to this reviewer. The authors claim that their work engages "in the provision of realistic and informed advice to clients in the sports sector".[2] This may seem to some to be an unrealistic separation. Some explanation should be afforded as to the nature of Lewis and Taylor's work: the book is essentially a collection of chapters relating to the various aspects of legal practice engaged by sports related activity. The edited edition contains contributions from many of the most eminent legal names associated with sport; drawn from what the editors suppose are the alien dominions of practice and academia. *Sport: Law and Practice* is organised around five broad themes: regulation, EC law, corporate, commercial aspects and the sports person. Although it is unwise to label any work comprehensive, the book provides an excellent resource for anyone interested in almost every aspect of the sport/law overlap. As a point of initial reference when considering a question of law pertaining to sport Lewis and Taylor's contribution is without parallel. Nevertheless, the reviewer must express a degree of perplexity as to exactly why it is that the editors consider the sports law/sport and the law debate to be of no relevance. Perhaps part of the reason is that the way in which the book is organised does not take particular account of whether a particular aspect of legal intervention pertains to the organisation of or participation in sport or to some peripheral activity; this seems crucial. In fact, the contributors to the collection do appear to consider the question relevant, at least implicitly; indeed in many areas it can be seen to be *the* crucial issue. In many respects the contributors are essentially asking the same question: is sport special? Should the law therefore accord it different treatment?

One does not have to delve very far into the work before these issues become wholly evident. Chapter A3 of *Sport: Law and Practice* considers legal supervision of sports

[1] At p vii.
[2] *Ibid.*

governing bodies. Interestingly, the author reasserts the view that the question of sports law is of solely academic interest,[3] which seems strange in a chapter considering cases in which the sporting context can be the only explanation for what would otherwise be considered perverse decisions. In *Stevenage Borough Football Club Ltd v Football League Ltd*[4] the High Court reversed the burden of proof in respect of the restraint of trade doctrine in order to protect the regulatory autonomy of the Football League; in *Finnigan v New Zealand RFU*[5] the New Zealand Court of Appeal stretched the contractual nexus to breaking point and beyond in order to be able to scrutinise a decision to send an All-Black rugby side to the then apartheid South Africa; and in *Agar v Hyde*[6] the Australian High Court declined to extend liability in negligence for faulty rule making which, it was argued, had caused injury, on the basis that the relationship between the sports body and those subject to the rules was insufficiently proximate. Similar examples can be found scattered liberally through the book. It would seem prudent for any person wishing to provide "realistic and informed advice to clients in the sports sector"[7] to have a grasp of quite why these courts felt it necessary to treat these cases differently, not just in the narrow doctrinal context but with an eye to a wider standpoint. A clear appreciation of judicial attitudes towards sport generally would undoubtedly benefit any practitioner in advising their client. Such an approach can only be described as "sports law".

This type of analysis is clearly evident in the work of Richard Parrish. In academic circles at least, Parrish must be considered to be amongst the pre-eminent authors in the field of European Community law and sport. For many years this mantle undoubtedly belonged almost exclusively to Stephen Weatherill, but Parrish's work *Sports law and policy in the European Union* provides a significant addition to a relatively neglected area. Unlike Lewis and Taylor, Parrish perceives that there is a great deal of value to be derived from a discussion of a "sports law jurisprudence". Very early on in the piece, he acknowledges Michael Beloff QC's warning of the inherent dangers attendant in sports law of description rather than analysis. In avoiding this approach, Parrish avoids falling into the main trap for the unwary venturing into the world of sports law: he avoids simply dealing with sport through consideration of its treatment by traditionally delineated legal sub-discipline. Parrish adopts a much more holistic appreciation of the approach to sport in Community law. His view is not that the law should be considered from the standpoint of a relatively dry doctrinal analysis, but that a more sectoral view is appropriate. In this respect, perhaps Parrish is assisted by the relatively narrow nature of Community law. EC law comes from a clearly economic position, and it may well be that such an approach is of assistance to the author. Parrish's work is strongly oriented around a desire to identify the rationale which underpins the approach of the EU towards sporting activity, in both its legal and political aspects. This much is clear from the outset, as Parrish begins by establishing the genesis of EU sports law and policy, with a view to establishing a coherent basis upon which it is possible to understand the interaction between sport and the EU. In particular, Parrish seeks to understand the different priorities of the various actors in this area and the need for the establishment of compromise in order to move forward. It is not until chapter four of the text that *Sports Law and Policy in the European Union* directly considers the role of the

[3] At p 88.
[4] [1997] 9 Admin LR 109.
[5] [1985] 1 NZLR 159.
[6] [2000] HCA 41.
[7] *Supra*, note 1.

European Court of Justice. This is significant. It is an acknowledgement that, although much of the high profile interaction between EU law and sport occurs before the ECJ, in reality the Court only serves to reflect a broader polity emerging within the institutions of the Union. It is only after this that Parrish does, indeed, consider a discrete area of law, the specific application of competition law in sports cases. While this might seem to follow the more narrow doctrinal analysis, the limitations of which have already been outlined, this is a necessary undertaking. Since the initial refusal of the ECJ in *Bosman*[8] to subject sport to competition law scrutiny, the area has rapidly developed into the most significant area of law relevant to sport. The application of competition law has also seen the maturing of EU law's approach to sport and many of the concepts developed have significance for other areas of law. In reality the evolution of competition law approaches to sport has done much to inform the development of the free movement of persons principles in this area, to the extent that it is arguable that the two are inextricably linked.

Parrish's work is undoubtedly the most important in this area and has a significance, not only in academic circles, but also to those who would wish to give sound practical advice to their clients in the sporting sphere. Crucially, Parrish does not stop at an analysis of what the law *is*, but also asks why this is the case and where the law is going.

Despite their protestations to the contrary, Lewis and Taylor's book does undertake a similar analysis in parts, even if this is not always as overt. Perhaps this is founded in the broad nature of their text. As suggested, *Sport: Law and Practice* is almost exhaustive in its coverage of the meeting of law and sport. Perhaps the problem is a failure fully to appreciate the importance of the sporting context in the application of the law. Because of the full range of sports related legal issues considered by Lewis and Taylor, it seems that the importance of the "sports law" debate is often lost.

The answer that seems to emerge from both Lewis and Taylor and Parrish's work, is that if the sporting context of a legal issue assumes significance in the outcome of a given situation then we must be talking about "sports law". Because Parrish deals almost exclusively with this type of situation, it is perhaps easier for him to discern this. But it is equally important for any practitioner to appreciate the consequence of the existence of a strong sporting context. It would be churlish to criticise *Sport: Law and Practice* too strongly in this respect; in fact much of this discussion is implicit in its legal analysis and the book is undoubtedly the best resource available concerning legal aspects of sporting activity. Equally, Parrish's work is of significance in that it represents a noteworthy development in the gaining of a greater understanding of the interface between sport and law.

SIMON BOYES*

[8] C-415/93 *Union Royale Belge de Sociétés de Football v Bosman*, [1995] ECR I-4921.

* Senior Lecturer, Nottingham Law School, Nottingham Trent University.

## LAW AND THE INTERNET

*Law on the Web: a Guide for Students and Practitioners,* by STUART STEIN, Harlow, Pearson Education, 2003, xi + 190pp, Paperback, £14·99, ISBN 0-13-060571-9

This review has been a long time in coming, mainly because it took me a very long time actually to read the book. I am not a particularly slow reader, and the book is a mere 190 pages long, but on numerous occasions I began reading, only to close the book midway through Part One. You may be developing the impression by now that this is not likely to be an entirely favourable review and I am afraid that that is the case.

The main criticism I have of the book is that it is not exactly clear what it is trying to be and, in this confusion, it suffers from trying to be too much. With the title *Law on the Web: a Guide for Students and Practitioners*, I was expecting a book substantially about law resources available over the internet, but descriptions of such law resources only occupy Part Three of the book. However, I am jumping ahead of myself, so let me deal with each part in turn.

Part One is entitled *Web Resources* and features three chapters, which cover an introductory glossary, understanding a URL and evaluating and referencing internet resources. Although there is undoubtedly useful information in here, such as how to decipher URLs and how to assess the quality and authority of a web page, I found the text too dense to be easily accessible. Some of the sentences required me to re-read them in order to understand the full meaning, even though I would consider myself a competent internet user. The "readability" of the text is also hampered by the inclusion of so many URLs and other examples of computer file names although, given the nature of the subject matter, it is hard to see any other way around this issue.

Part Two, *Locating Internet Resources*, covers internet subject directories, internet gateways, search engines and mailing lists. Again much of the information is useful, but too wordy and I found myself wishing for more graphical examples and less text. Leaving my personal preference for visual stimulus to one side, I am sure that there are many useful advanced searching tips here for law students. What I am less sure about is how many law students looking for tips on searching Google would reach for a book called *Law on the Web*. Similarly, while the information on locating and joining mailing lists is quite detailed, I cannot help feeling students would be more likely to reach for a book about mailing lists for this advice.

Part Three is entitled *Law Internet Resources*: things are looking up. It may have taken until page 67, but now I have arrived at the law subject directory. The focus here is on publicly accessible web pages, so no subscription databases are included, but the directory does list a large number of sites covering a range of different subjects. There are broad subject headings, such as case law/legislation, associations/institutes and legal systems and also narrower headings such as asylum/immigration, employment law and human rights. Some of the headings (and contents) are slightly surprising, namely travel, which contains URLs for cheap flight websites and an online air distance calculator. URLs are provided for each resource listed and most also contain some description of the resource, although not all. One of the real plus points is that the author's website (URL available in the book) contains a list of all the URLs referred to, so they could be saved and used to create a ready-made favourites list. Although the preface states that the publisher's web site will be amended when URLs change, I was disappointed to find some links which had not been updated. While I appreciate (as the author points out) that URLs can change from one day to the next, I was aware that the URL for a major library catalogue had actually changed many months ago, but had yet to be amended on the author's list.

Part Four is, *Sundry Matters* and I am confused again, as these chapters cover compiling web pages, miscellaneous computing matters and browser features and customisation. Again, I have to ask myself if this is a suitable place to try to teach people about hypertext mark-up language, page layout or how to market your web pages? And, again, I come to the conclusion that people who want to know about those things, or indeed how to download and install software, would probably look in a computing book rather than this one.

So, it probably comes as no surprise to say that I was disappointed overall by this book. To call it *Law on the Web* seems misleading, as so little of the content is specifically about law. The law subject directory in Part Three is very useful and I am sure that I will use many of the links provided *via* the publisher's web site, but if I wanted information about writing web pages or the internet in general, I would be looking for a book with a lighter writing style and more visual examples. A case of "jack of all trades"...?

ANGELA DONALDSON*

* Information Specialist (Law), Nottingham Trent University.

NOTTINGHAM LAW JOURNAL

VOL 14(2) 2005

# NOTTINGHAM MATTERS

*This section documents major developments and research projects within Nottingham Law School together with responses to public consultation exercises and other public contributions made by its staff.*

## SPORTS LAW AT NOTTINGHAM LAW SCHOOL

The growth of sports law as a discrete legal discipline is not uncontroversial; indeed much of the most heated debate in the area concerns this very question.[1] The topic is one that forms an important part of growing range of provision at Nottingham Law School.[2]

### ACADEMIC PROGRAMMES

During the academic year 2004/5 Nottingham Law School has initiated two major developments in sports law. At undergraduate level, the School now offers final year law students the opportunity to study a module in sports law and regulation, which provides an introduction to legal and regulatory issues pertaining to aspects of sporting activity. These are as diverse as the criminal and civil law aspects of on-field violence; liability for sports injuries; combat sports; disciplinary proceedings; doping and drugs; European Community law; competition law; media rights; venue safety and spectator disorder. The module begins by challenging students as to the existence of "sports law", with the intention that they will be able to present an intellectually satisfying account by the end of the course.

In parallel with this undergraduate provision, Nottingham Law School now offers postgraduate students the opportunity to study towards a Master of Laws (LLM) qualification in sports law. The programme offers specialist modules in a range of related areas. The core module – *Sport, Law and Regulation* – introduces students to the key concepts in relation to the law-sports relationship and establishes the contextual framework around which the programme operates. The *Sport, Law and the Human Body* module considers the legal relationship between participants, officials, coaches and rule-makers as well as medico-legal aspects such as treatment for injuries; whilst in *Sport, Law and Business* commercial relationships in the sporting arena are scrutinised. These modules are all undertaken between October and the Christmas vacation alongside specialist training in legal research and writing. In the New Year students undertake four specialist modules: *European Sports Law and Policy* provides

[1] See the book review in this journal.

[2] Nottingham Law School also offers the following public LLM programmes: Corporate; Europe and the Law; General (by selection of options); Health; International Criminal Justice; International Trade; and, for practitioners, Intellectual Property Litigation; Commercial Intellectual Property; Advanced Litigation and Dispute Resolution as well as an MBA in Legal Practice.

a critical insight into the crucial and shifting interface between EU law and sport; *Intellectual Property Law* looks at the opportunities and problems in respect of intellectual property aspects of sports statistics and information, image rights, events rights, endorsements, and sports goods and equipment; *Sport and Employment Law* considers the variety of employment relationships encountered in the area and the peculiarities which are thrown up; in *Sport, Law and Society* the legal issues pertaining spectator safety and sports related violence are considered. On successful completion of this programme of study students then complete a substantial dissertation on a topic of their choice. Those studying towards the Law School's other academic LLM programmes can also choose sports law modules in their programme of studies.

The Sports Law LLM programme is supported by Freeth Cartwright LLP, a leading regional firm of solicitors, extremely active in the field, which offers a prize for the best student.

## ACADEMIC EXPERTISE

The delivery of sports law teaching at both undergraduate and postgraduate level is undertaken by a wide range of staff. Simon Boyes is heavily involved in both programmes. He was previously connected to the Sports Law Centre at Anglia Polytechnic University and the Centre for Studies in Sports Law at King's College, University of London. Simon co-authors the leading student textbook in the area and has written extensively on the subject.[3] Paul Lloyd Davies combines legal expertise with actual participation in a number of combat sports and martial arts. Other members of the team bring specific expertise to bear in the context of sports law. Austen Garwood-Gowers is an expert in legal issues pertaining to the human body[4] and has also played cricket to a high level. Tom Lewis is a specialist in public order and human rights,[5] as well as being a keen rock climber. As well as leading the academic LLM programmes, Kay Wheat contributes a wealth of knowledge in respect of venue safety and employment law.[6] Ewan MacIntyre writes one of the leading textbooks on business law[7] and is a keen sports fan. Peter McTigue has recently joined the Law School, having been a practising solicitor with many years of experience in the field of employment law. The broader interest in the topic is highlighted by the longstanding interest of the Dean of the Law School, Professor Michael Gunn, who has published widely in respect of criminal law aspects of sport.[8] The LLM programme also draws on expertise from elsewhere in the university to ensure a well-rounded student experience.

[3] Gardiner, S; James, M; O'Leary, J; Welch, R; Blackshaw, I; Boyes, S and Caiger, A, *Sports Law* (2005: 3ed, London, Cavendish).

[4] See *eg*, "Improving protection against indirect interference with the use and enjoyment of home: challenging the legacy of *Hunter v Canary Wharf* using the European Convention on Human Rights and the Human Rights Act 1998", (2002) 11(1) *Nott LJ* 1.

[5] See *eg*, "Democracy, free speech and TV: the case of the BBC and the ProLife Alliance" (2005) 3 EHRLR. 290; "Human earrings, human rights and public decency" (2004) 5 Web JCLI.

[6] See *eg*, "Proximity and Nervous Shock", (2003) 32(4) CLWR 313; (2002) 65(3) MLR 425; *Napier and Wheat's Recovering Damages for Psychiatric Injury* (2002, Oxford, OUP).

[7] MacIntyre, E, *Business Law* (2005, Harlow, Pearson).

[8] Gunn, M and Ormerod, D, "The legality of boxing", (1995) 15(2) LS 181; Gunn, M, "The impact of the law on sport with specific reference to the way sport is played" (1998) 3(4) CIL 221; Gunn, M and Ormerod, D, "Despite the Law: Prize-fighting and Professional Boxing" in Greenfield, S and Osbourne, G, (eds) *Law and Sport in Contemporary Society* (2000, London: Frank Cass), p 21. See also Huxley, P, "The boy done great: football and free movement", (1996) 5(1) *Nott LJ* 97.

## STUDENT SUCCESS

Engagement with the sports law project by the student body has been excellent. Both undergraduate and postgraduate students have been successful in obtaining prestigious placements with leading law firms and sports bodies such as the British Olympic Association and UK Swimming. A number of the initial LLM cohort has been successful in securing sports-related posts even before completion of the course. In particular, one student has secured a role with a leading European sports marketing company and another with a major English county cricket club. Other members of the course have had scholarly success, having been published in academic journals and involved in collaborative research projects with Law School staff.[9]

Former students have continued this success: Elizabeth Emery won the Lord Woolf Scholarship for 2005 by writing an essay discussing the rationale for legal intervention in sport and her winning submission was later published in the *Daily Telegraph*.[10]

## FUTURE PROSPECTS

The genesis of the sports law project at Nottingham Law School has undoubtedly been a success; one on which the School will be looking to build through developments to academic provision and advancing expertise in the area, as well as partnership with the legal profession and sports organisations.

SIMON BOYES*

[9] *eg*, Patel, S and Boyes, S, "Football: Legal and Regulatory Approaches to Women in a 'Men's Game'": paper presented at the International Football Institute Conference, *Women, Football and Europe*, 12–15 June 2005.

[10] *Daily Telegraph*, 31 March 2005.

* Senior Lecturer, Nottingham Law School, Nottingham Trent University.

# TABLE OF CASES

# TABLE OF STATUTES

# TABLE OF STATUTORY INSTRUMENTS

# TABLE OF EU LEGISLATION