

CONTENTS

Personal Pensions and The Financial Services Authority: New Chapter or Same Old Story? (Marina Milner and Keith Syrett) 141

Past Human Rights Violations - The MacDermott Lecture (Mr Justice Richard J Goldstone) 164

Disadvantage and Discrimination: The Emerging Jurisprudence of the South African Constitutional Court (Evadne Grant & Joan G Small) 174

Of Lamé Ducks, Black Sheep and Family Bonding (Sheena Grattan)..... 198

Modernising the Regulation of Water Pollution in Northern Ireland - Part II (Sharon Turner) 230

Grant-ing Rights: the Politics of Rights, Sexuality and European Union (Carl F Stychin) 281

The Model of Social Relations: Strategic Behaviour, Trust and Loyalty (Joseph Savirimuthu) 303

Anonymous Witnesses (Ruth Costigan & Professor Philip A Thomas) 326

Published four times yearly by SLS Legal Publications (NI), School of Law,
The Queen's University of Belfast, Belfast BT7 1NN, Northern Ireland.

ISSN 0029-3105



PERSONAL PENSIONS AND THE FINANCIAL SERVICES AUTHORITY: NEW CHAPTER OR SAME OLD STORY?*

Marina Milner and Keith Syrett, Norwich Law School, University of East Anglia

INTRODUCTION

Despite origins embedded within Thatcherite anti-welfarist policies and ideologies of individual choice, the personal pension (PP) scheme appears to have become a permanent feature of the United Kingdom's pensions landscape.¹ However, the revelation that large numbers of such schemes were 'mis-sold' to persons for whom they were not the most suitable option has prompted demands for "better regulation to restore confidence in the system".² Against such a backdrop, the Labour Government has embarked upon a radical overhaul of the entire structure of financial services regulation, including that of PPs. This article seeks to assess the suitability of the new regulatory regime in the PP context, in view of the particular qualities that serve to distinguish this product sharply from others in the financial services marketplace.

PENSIONS MIS-SELLING: A BRIEF BACKGROUND

PPs are regulated as emanations of the financial services industry. As such, under the Financial Services Act 1986, they were subject to the powers of the Securities and Investments Board (SIB) and the relevant self-regulatory organisations (SROs). It is noteworthy that at no time was any single SRO responsible for the regulation of PPs *per se*; instead, tasks were divided along functional lines. When evidence of mis-selling began to emerge, the key SROs were Fimbra, which regulated Independent Financial Advisers (IFAs), and Lautro, responsible for regulating life offices. Between 1990 and 1992, Lautro and Fimbra began to receive information that suggested low levels of compliance with regulatory standards. In July 1992, the SROs therefore issued more particular guidance as to what was required,³ and the following year SIB commissioned a report by KPMG Peat Marwick to investigate compliance with these new guidelines.

The report revealed that those who had been in occupational pension (OP) schemes were being advised to transfer to PPs, even though these were less

* This article should have appeared in the Spring Issue of the *NILQ*. The article describes the position at the date of acceptance, *i.e.* December 1999 and SLS apologises for the delay in publication.

¹ Current proposals for pensions reform envisage a growing role for private provision, including PPs: see DSS, *A new contract for welfare: partnership in pensions* (1998; Cm 4179): *passim*.

² *Ibid.*, Summary, para 34.

³ See now SIB, *Pension Transfers and Opt-Outs – Further Safeguards for Future Business Trends* (March 1994); *Pension Transfers and Opt-Outs: Model Guidance* (May 1994) *etc.*

suitable because (for example) they did not receive employer contributions and did not guarantee a salary-related level of income on retirement. Similarly, those who could have joined an occupational scheme did not do so, on advice, and instead bought PPs. The report referred to “widespread regulatory compliance failure”⁴; the cost to providers and financial advisers of making good the mistakes made has since been estimated variously at £4 billion to £11 billion.⁵

The reasons for such large-scale breach have been analysed fully elsewhere.⁶ It is submitted here that pension provision is a unique activity and that the PP is unlike any other form of financial service. Regulation must identify and manage the particular risks that continue to arise: the failure of the previous regime to do so may have contributed to its failure. The prospects for success of the new regulator and our proposals for the future will be considered below. First, however, the uniqueness of pension provision and PPs as manifestations of such provision will be explored.

THE DISTINCTIVE QUALITIES OF PERSONAL PENSIONS

1. Social risk and pension provision

In many senses, pension schemes are merely savings schemes – a type of financial service. Contributing to them or providing them is therefore an unremarkable economic activity. As an analysis of pension provision, however, this approach is unsatisfactory because it does not take account of the distinctive factors at issue. Pension provision caters for the risk of reaching old age without income. Unlike other forms of social risk, where the degree of risk is more uncertain, most people are more likely than not to grow old. In addition, the social convention of retirement from the job market, which is now so well-established as to be a norm, means that most will have to deal with the issue of how to support themselves during retirement. Although some will be well placed to do this, the risk of a “penurious old age”⁷ is much more widespread than other similar social risks. Because it is so universal, the costs of dealing with it are huge: in 1990, the combined expenditure on public and private pension schemes in the United Kingdom was approximately 11% of GDP.⁸ Government literally cannot afford not to deal with this form of risk, but its current income is not sufficient to underwrite it.

⁴ KPMG Report (1993), p 16.

⁵ See *Occupational Pensions* (March 1998), p 3 and Treasury Select Committee, Ninth Report, *The Mis-Selling of Personal Pensions* HC 712-I (1997-8), p x. There are no figures available enabling an estimate to be made of the numbers of PPs mis-sold in Northern Ireland: see 299 HC Deb 160 (written answer, 29 July 1997). However, the FSA has indicated its awareness that a problem exists (communication with authors).

⁶ See generally KPMG Report, *op cit* n 4; Black and Nobles, “Personal Pensions Mis-selling: The Causes and Lessons of Regulatory Failure” (1998) 61 *MLR* 789 at 793; and for matters following the pensions review see Treasury Select Committee, *ibid*.

⁷ Baldwin, *The Politics of Social Security* (1980), p 12.

⁸ Barr and Whynes, *Current Issues in the Economics of Welfare* (1993), p 45.

There is also another factor. It is well documented that the population is ageing, so that dependency ratios are falling: in 1953, there were 4.6 people of working age for every pensioner, which compares with 3.4 today and an estimated 2.4 by 2040.⁹ It will therefore become increasingly unrealistic to rely on current earners to support all non-earners.¹⁰ Nevertheless, assuming that the current philosophy of welfare survives, those who grow old will still need support. Plans are needed to stave off what could be a major economic problem.

The breadth of the risk faced and its financial consequences are not, however, the only factors that make this a unique issue for social policy. When ideas about social justice gained enough weight to be translated into government policy at the turn of the century, two elements on which the claim of the elderly to state support rested emerged, which made their demands very difficult to resist. Previously, the elderly had been treated as a subset of the poor and therefore subject to the same rules with regard to poor relief as others. However, there was a growing recognition of the destitution suffered by many of the elderly through poverty. Coupled with this was the fact that, unlike many other forms of social need, old age was often accompanied by undoubted desert.¹¹ Although some individuals may have wasted away their lives in 'feckless poverty' (as proponents of the Poor Laws may have put it), many more seemed to be suffering poverty despite having worked all their lives. Thus, those claiming relief in old age could point to two justifications: need *and* desert – which otherwise have often been set in opposition to each other.¹² In this respect they were perhaps unique amongst those claiming social support.

Partly because the risks faced are so great and expensive, therefore, and partly because, politically, the twin elements of need and desert are still apparent and powerful¹³, old age, retirement generally is now a distinct

⁹ DSS, *op cit* n 1, ch.1, para 27. Approximately 10 million people currently receive state pensions; by 2030, the number will rise to about 14 million: see Pension Provision Group, *We all need pensions – the prospects for pension provision* (1998), Part 2.

¹⁰ See further Miles and Timmermann, "Risk sharing and transition costs in the reform of pension systems in Europe" (1999) 29 *Economic Policy* 253 at 257. Note, however, that several commentators have cast doubt on the negative impact of the so-called 'demographic timebomb': see Vincent, "Who's Afraid of an Ageing Population?" (1993) 16 *Critical Social Policy* 3 at 23; Hills, *The Future of Welfare: A Guide to the Debate* (1993), pp 12-13; Barr and Whyne *op cit* n 8, p 50 *et seq.*

¹¹ For contemporaneous accounts and acknowledgement of need, as well as attitudes towards 'desert', see *eg* Spender, *The State and Pensions in Old Age* (1892), ch.1; and generally Committee on Old-Age Pensions, *Old Age Pensions: A Collection of Short Papers* (1903).

¹² See *eg* Baldwin, *op cit* n 7, p 31 *et seq.*; and generally Miller, *Social Justice* (1976). For the history of the development of old age relief, see *eg* Brown, *Social Security for Retirement* (1990), pp 3-47; and for insurance as a model for welfare provision, see Ogus *et al*, *The Law of Social Security* (1995), p 25 *et seq.*

¹³ See *eg* Pinker, *Social Work in an Enterprise Society* (1990), p 131, drawing on data from public opinion surveys. He notes the wide support for state health, education and pensions provision as compared with the unemployed; and comments on the implication that the former are 'deserving' groups. It is also

socio-political issue, and it has become necessary for government to make some provision for pensions. The cost element involved has meant that the choice of funding method has been an issue since state provision began.¹⁴ One solution is to adopt the ‘savings model’ of provision, where an individual saves income generated whilst economically active, so as to provide assets on which he can live in retirement. In this way, income can be drawn to support the current retired population “without any impact on the [current] working population”¹⁵. Politically, one of the crucial implications of this is the focus it places on the individual and current earnings, and the consequent non-involvement of the state and society either in provision itself or, in principle, in any failure of the individual’s arrangements. Although it may be accepted that a certain level of state welfare provision is unavoidable (most western industrialised states now provide it at some level), if the current workforce is not involved in the support of its predecessors, then responsibility falls on the individual in each workforce to be self-reliant in retirement. This approach assumes that the individual can earn enough during a working life to amass adequate savings for retirement and implies that self-reliance and the effectiveness of the individual to provide for himself are paramount.

This can be contrasted with the so-called ‘solidarity model’.¹⁶ As its name suggests, this involves a measure of collective action and responsibility for the welfare of all, born out of a sense of community within society¹⁷, it may be described as the acquisition by the ‘retiree’ of a claim to output by the next generation of workers.¹⁸ Thus, if individuals fall foul of risk, society has a responsibility to help bear the cost consequences. Typically, social risk has been met through some form of re-distributive insurance¹⁹. This functions by recognising that there are certain risk pools into which more than one person might fall. Given this, the adverse consequences of those risks manifesting themselves can be minimised if all those at risk contribute money to a common fund, from which a cash sum can be paid to those for whom the risk does materialise. Within the private sector, this means “reapportioning the costs of risk and mischance”²⁰ amongst those in the risk pool (rather than, for

interesting that the keynote of the Government’s reform for welfare as a whole is that “work is the surest route out of poverty”: see DSS, *New ambitions for our country: a new contract for welfare* (1998; Cm 3805), Summary, para 10, and generally.

¹⁴ In 1908. See eg Brown, *op cit* n 12, p 3 *et seq.*

¹⁵ Davies, *Better Pensions for All* (1993), p 7; and see further Barr and Whynes, *op cit* n 8, p 48 *et seq.*

¹⁶ Davies, *ibid.*

¹⁷ An example of the adoption of this approach in state policy is contained in *Social Insurance and Allied Services* (‘the Beveridge Report’, 1942; Cmd. 6404), para 26. In dealing with the eradication of ‘Want’ which affected *inter alios* the elderly, Beveridge aimed neither to raise living standards, nor to reduce inequality, but to “enhance a sense of collective belonging within society”: Langan, *Welfare, Needs, Rights and Risks* (1998), p 8.

¹⁸ Barr and Whynes, *op cit* n 8, p 48.

¹⁹ This satisfies the ‘desert’ element of the claim to social support. See generally de Vries, “From Insurance State to Welfare State, and Back Again?” (1995) 20 *Geneva Papers on Risk and Insurance* 439.

²⁰ Baldwin, *op cit* n 7, p 1.

example, re-distributing wealth *per se*). When social insurance is in question, two extra steps are taken. First, those who are not so clearly within the risk pool may contribute to the costs of insuring all who do fall within it; second, the cost of contributing to the fund may be re-apportioned a second time, this time according to ability to pay.

What is at issue here is not the justice or otherwise of the wealth redistribution aspects of social insurance as opposed to its risk re-apportionment. It is simply to identify the characteristics of the vehicle chosen to meet, and defray the consequences of, the particular social risk. The PP appears to be a manifestation of the 'savings model' alone²¹: it is tied specifically to the individual and its effectiveness for retirement provision depends on how much that individual sets aside during a working life. Although the basic state pension would currently provide some back-up cover if the model fails (resulting in inadequate pension provision), this is not linked *per se* to the creation and development of the PP. Hence, within the product itself there is no element of 'solidarity': there is no element of risk-spreading. Additional factors relating to the PP itself are that even if the most appropriate product is sold, investments may still fail, or return otherwise undermined because, say, of unfavourable annuity rates at the point of retirement.²² Since there is no security of return, it is possible for a buyer to draw out less than was paid in, no matter how much that might be.

Once again, the point is not to make a value judgement about the choice of funding method, but rather to analyse the effectiveness of the 'savings model' as a means of achieving the given end: the provision of income upon retirement. The equation of pension provision with finance through the PP has placed it in an environment (financial services) where the social and individual goals of provision may not be met because the means of funding exposes the buyer to unrecognised or unregulated risks. As we have seen, there is no guarantee that the financial mechanism for providing the pension will do so (as there is, for example, in both occupational and state pension

²¹ John Denham MP described these second pensions as "solely dependent on each individual's ability to work and save": speech to Northern Pensions Conference, "A Third Way for Pensions Reform" (10 July 1998). See also Miles and Timmermann, *op cit* n 10 at 254, who observe that the PP is a "pure defined contribution scheme. . . with no redistribution between or within cohorts".

²² Annuities provide income for life in return for the investment of a lump sum. Legally, those retiring with PPs or OPs must buy an annuity by age 75. But annuity yields have dropped dramatically in recent years, with the loss of approximately £5000pa income for every £100,000 invested. Some insurance companies guaranteed annuity rates in the 1970s and 1980s, when it was difficult to imagine that yields would drop so much; the cost to them of meeting these obligations has been estimated at £14 billion (see *The Independent* 3 March 1999, *The Financial Times* 14 August 1999). Although some companies appear to be honouring these promises (see *The Independent* 9 April 1999), the Equitable Life mutual insurance company has recently resisted a legal challenge to its policy to honour its guaranteed rates over only a small proportion (or none) of the fund values it invested on behalf of its members (so disappointing the expectations of some of its members, though in order, it argued, to achieve parity between all classes of member). See *Equitable Life Assurance Society v Hyman*, *The Times* 12 October 1999; and press reports generally.

schemes)²³. Focusing solely on the capital markets means also that there is no cover for failure of individual provision, and no recognition of either desert or need.

The nature of the risks accompanying the PP is examined in further detail below. For the present, however, the problem and its repercussions might be expressed as follows. Whilst the origins of the discourse about pension provision may have been in welfare, the admitted practical difficulties of resourcing this specific aspect of welfare have meant that the focus of the debate has long since shifted to means of funding. Given the mechanism chosen, the dominant discourse surrounding the PP has become that of finance/investment.²⁴ The financial services market may generally be an effective means of raising funds: that is not in issue. But within that particular discursive community, certain matters are taken for granted that demand more cautious treatment when pension provision is the only aim of the venture in question. Risk is the prime example. Those ‘investing’ inevitably face risk; indeed, the argument can be made that finance is *about* risk. The finance/investment discourse therefore allows us to dismiss concerns about the PP and the risks it presents to buyers simply by pointing to the nature of the ‘game’ involved:

“It is not reasonable to be protected against all possibility of losing money, or of making mistakes. . . finance is necessarily about risk, and therefore to attempt to regulate to remove all risk from the consumer would be a policy of regulating away the very function of finance and financial contracts”.²⁵

On this analysis, the most that regulation can achieve is to prevent people being ‘made fools of’ in financial terms²⁶.

This surely misses a step. Pension provision *per se* is not about risk (except in the sense of managing social risk) and certainly not in the same way that finance/investment is. Salary-related OP schemes are examples of how the risks *of investment* are managed so that they do not fall on the intended ‘retiree’; hence such schemes have a maximum chance of achieving their goal.²⁷ Indeed, trustees who invest funds for OP schemes are subject to strict trust law rules regarding attitude to and levels of risk that are acceptable.²⁸ We assume here that the overriding objective is to achieve the distinct social policy aim of providing adequate income on retirement. Given this, the risks

²³ Davies argues that “half the time people [buying PPs] will end up with a pension that is less than they were led to expect”, *op cit* n 15, p 28.

²⁴ The most notable example being the definition of the PP as an ‘investment’ in the Financial Services Act 1986, Sched. 1, para 10.

²⁵ Llewellyn, “Regulation of Retail Investment Services” [1995:Spring] *Economic Affairs* 12 at 16; also FSA, *Meeting our Responsibilities* (1998), para 47: “no system of regulation can insulate even retail customers from the responsibility of taking their own decisions on their savings and investments”.

²⁶ Gower, *Review of Investor Protection*, Part I (1984; Cm 9125), para 1.16.

²⁷ Other types of risk may, of course, arise *eg* from fraudulent employers as in the Maxwell case.

²⁸ See *Re Whiteley* (1886) 33 Ch D 347 (CA): “the duty of a trustee is. . . to take such care as an ordinary prudent man would take if he were minded to make an investment for the benefit of other people whom he felt morally bound to provide”, *per* Lindley LJ at 355.

inherent in the PP and the financial markets should at least be overtly recognised and managed, through regulation. The nature of these risks will next be examined.

2. An environment of risk

The PP involves numerous forms of risk. The first encountered are those inherent in making any form of pension choice. The mis-selling saga highlights some of these: an individual considering pension provision may, if employed, have to consider whether to join an OP scheme rather than taking out a PP, or to opt out of one in favour of a PP. Many find this a confusing choice,²⁹ while others may make it on non-financial grounds;³⁰ but those in secure employment with access to a good OP scheme may be exposed to a higher risk of receiving less on retirement if they buy a PP because PPs generally do not attract employer contributions and do attract fees and commission. Even given a financially knowledgeable decision – and even if an employer makes contributions to a PP – factors such as investment performance, the structure of commission and fees charged, and annuity rates at the point of retirement may all combine to make the PP less successful than the OP.³¹

Quite apart from the risks inherent in choosing between OPs and PPs, as illustrated by mis-selling, there are investment risks inherent in any type of funded scheme. The conclusions of Miles and Timmermann illustrate this. They have examined the relative efficiency of funded and unfunded provision by governments across Europe and note that accepted wisdom appears to understate “significantly. . . the true volatility of the terminal value of portfolios”. They find that the number of recessions that are experienced during the “active work and savings period” of a particular age cohort of the population cannot be evened out over time, despite the commonly held belief that the contrary is true, and conclude that “even in the long run, the risk of funded pensions in the stock market can be considerable”.³² In the light of the findings that all funded schemes involve risk, those inherent in PPs may be viewed as particularly significant, given the absence of any possibility of risk-sharing in such vehicles.³³ These are important points when considering the suitability of the PP for the social goal of pension provision.

Some types of risk have already been recognised and addressed by regulatory authorities in response to the mis-selling episode.³⁴ However, other forms of risk remain as inevitable consequences of the PP’s location within the markets and within the discourse of financial speculation. Here the risk is that investments can fail as well as succeed. Varying degrees of such risk exist: at one level, there is a situation approaching chaos, or maximum risk of

²⁹ See *eg* Pensions Law Reform Committee Report (‘the Goode Report’, 1993; Cm 2342), vol. II, Research Report 2 (generally).

³⁰ Because, for example, of distrust of the employer: see Goode Report, *ibid.*, at 135.

³¹ *I.e.* because the return from the PP may be less than the pre-defined income guaranteed by formula under a salary-related OP

³² Miles and Timmermann, *op cit* n 10 at 266, 254, 270, respectively.

³³ See above, n 21 and accompanying text.

³⁴ Above n 3.

financial loss. No distinction is made between the individual and the corporate investor: each is responsible for fact-finding and choice of investment or provider, and the investor who chooses well can make large profits.³⁵ The opposite condition is one of complete security for the individual: he is guaranteed not to lose money and in some situations will be guaranteed a certain return. This will be important if he is relying on that return to live on, having no other means of support; however, profits are likely to be relatively small and investment return relatively slow.

These risks are confronted by a 'spectrum' of investors/consumers.³⁶ The characteristics of each individual will include rational capacity, wealth (disposable cash or otherwise) and attitude to risk. The PP purchaser might, as part of any of these characteristics or in addition, also be affected by the level of need for income on retirement. These attributes can combine in several ways to produce a consumer who is 'vulnerable' to varying degrees, ranging from the investor with plenty of disposable cash and who is happy to take any kind of risk, however great, to an individual (such as our PP consumer) purchasing with the specific purpose of living on the funds raised because no other income will be available at that time.

Various factors will determine where any individual will sit within the 'risk environment', not all of which will necessarily foster the desired outcome: that is, to ensure that the retirement provision vehicle chosen is in fact likely to provide adequate income in retirement. One such factor relates to the nature of the market involved: wholesale, that is, investments commonly dealt in by professional investors; or retail, where consumers are involved. Focusing solely on the retail market, the nature and level of sophistication of the individual next becomes relevant. Black describes retail market consumers as, at one and the same time, "usually unsophisticated with a low level of understanding of the financial products they are buying", uninterested and unwilling to shop around or find out more; yet feeling capable of handling their own finances.³⁷ As she says, "[t]his dissonant

³⁵ See Pagnano, "Discussion" in Miles and Timmermann, *op cit* n 10 at 281: "as any student of capital markets knows – higher expected rates of return come at the cost of greater risk".

³⁶ The new regulator has argued for a three-way classification of consumers, distinguishing between professional traders, "expert end-users" and the man in the street: see Joint Committee on Financial Services and Markets, First Report, *Draft Financial Services and Markets Bill* HC 328-I (1998-9), para 28. The Government, however prefers to view the range of consumers/investors as a spectrum, so allowing for more subtle gradations between the categories: see below, n 79.

³⁷ Black, *Rules and Regulators* (1997), p 143, citing OFT study (1992). Other research shows that those with PPs (and members of OP schemes) were more likely than those receiving the State Earnings Related Pension to have thought "a lot" or "some" about pension provision in general and that a high proportion (about one half of the sample) felt that they had "very" or "fairly good" understanding of their schemes. The same survey reveals low levels of understanding about pension choices in general and the specifics of PP design: see Williams and Field, *Pension Choices, A Survey on personal pensions in comparison with other options*, DSS Research Report 22 (1993), paras 1.4 and 1.7 respectively. On the latter see also the Goode Report, Research Reports 1 and 2, *op cit* n 29.

combination of confidence and lack of sophistication suggests that consumers are especially vulnerable on savings and investment issues”.³⁸ She also points out the dangers of combining this type of consumer with a complex and opaque product such as the PP and the risk inherent in the competitive market.³⁹ Such difficulties are particularly acute for the poor, largely because of inequalities of status between users and providers, especially when one party has a “monopoly, or near monopoly, of expert knowledge”.⁴⁰ Nor is the provision of additional information necessarily helpful in redressing the balance, as consumers can simply be confused by it.⁴¹

Thus the PP market faces general difficulties in terms of competition, which increase the risk factors faced by our consumer. In addition, there are issues that specifically relate to the PP product. As Pinker states, the “ultimate sanction of the buyer [of such services/products]. . . is his ability to withdraw his custom altogether”.⁴² It is, however, difficult for the buyer of a PP to do so. For some, the PP replaces state provision for a socially recognised need (and it was promoted for this purpose).⁴³ In this situation, withdrawal is very difficult. Also, because of the nature of the PP mechanism, it is difficult for a consumer to know when it might be advantageous to withdraw. As Simpson argues, the future investment performance on which the buyer depends is something that can be estimated, but not known. In any event, withdrawing too early from the PP can be disastrous for an ends-oriented consumer: the commission due on some products is paid out of the early contributions made – say, over a period of two years. This means that no return will be made on the investment during that time. The ideal, therefore, may be for a PP contract to last for a long period, but this in turn means that the consumer does not gain experience of the product by frequent purchase. Nevertheless, the PP involves an ongoing financial commitment that may absorb a large

³⁸ Black *ibid*, citing research commissioned by SIB and Lautro (1991).

³⁹ Ogus *et al* make a similar, general comment, that the market model of welfare “assumes that individuals have the information available *and the capacity* to make rational, wise decisions concerning current and future welfare; unhappily, this is not always a realistic assumption” (emphasis added): *op cit* n 12, pp 9-10.

⁴⁰ Pinker, *Social Theory and Social Policy* (1979), p 143. He argues that the poor “lack the money to claim parity. . . Normatively and relationally they learn to define themselves as inferior persons, subordinate in terms of both money and knowledge”. The Pension Provision Group has observed that lower income groups are more vulnerable to the risk element that is inherent in PPs, *op cit* n 9, p 83. Davies has also commented on the impracticality and ineffectiveness of the PP as a means of pension provision for those on a low income, *op cit* n 15, pp 24-29. See further below, n 118.

⁴¹ The Consumers’ Association has commented that existing disclosure of information on charges “had not. . . resulted in a fall or increased competition, as consumers were ‘confused by this plethora of information’”: see Treasury Select Committee, *op cit* n 5, p ix.

⁴² Pinker, *op cit* n 40, p 144.

⁴³ This phenomenon has been described as ‘coerced exchange’. This is the idea that “people may effectively be ‘coerced’ into buying goods or services for which they previously depended upon state provision”, hence preventing a reduction in welfare rather than increasing it *per se*: see Dunleavy, “The Growth of Sectoral Cleavages and the Stabilization of State Expenditures” (1986) 4 *Society and Space* 129 at 143.

proportion of the consumer's wealth, which has corresponding implications for any mistakes made.⁴⁴

Whilst PPs therefore rely on the free market to provide an adequate source of pension provision, this ideal scenario is inherently skewed because the consumers will frequently not act appropriately for the market mechanism to work effectively. In terms of placing consumers within the 'risk environment', the factors identified in this section mean that the decisions they take are high-risk, because they will be less able to deal with the greater knowledge, power and sophistication of those selling a product which may often appear complex. Hence, they face a greater risk of losing money, and therefore of failing to secure adequate retirement income.

3. Regulatory consequences: the management of risk

Each of the risks identified above contributes to the danger that the PP product will 'fail' in its objective: the provision of retirement income. It is this objective, and the accompanying social and political consequences, that makes the PP distinct from other types of 'investment' in the financial markets.

We argue, therefore, that attention should concentrate on the central purpose of the PP, rather than the mechanism (financial services) by which it is funded. The goal is pension provision rather than investment, the point being "not to build up an assignable asset but rather to provide a means of fiscal security for the pensioner and his family".⁴⁵ The technical design of the PP supports this objective: in order to obtain tax exemptions, providers must incorporate a contractual prohibition on assignment.⁴⁶ It is designed for the sole purpose of security in retirement, and since the funds are not available until an agreed retirement age, individuals are unlikely to see it in any other light. Even if those who are better off buy PPs, their purpose will still be retirement income: a PP consumer who can 'afford to lose' the money contributed because he has other sources of income will still focus on the *ultimate outcome* of the investment, not on the investment activity itself. Its loss therefore signals the profound failure of the venture. This is especially so, the needier the individual concerned. The fact that, for example, NHS workers and miners were among the victims of mis-selling – a function of previous government policy that all levels of worker could purchase them – demonstrates the special dangers arising when those who genuinely need the income from PPs buy them.

⁴⁴ Simpson, *Regulating Pensions: Too Many Rules, Too Little Competition?* (1996), p 20. Again, the Treasury Select Committee comes to similar conclusions: *op cit* n 5, p ix.

⁴⁵ Thomas, "Bankruptcy in Occupational Pension Schemes and Personal Pension Schemes: A Dichotomy of Approach" [1998] *Conveyancer* 317 at 325. The National Consumer Council also appears to recognise this approach: "our view is that . . . the people who are buying [are doing so] really not as an expert purchase but as a means to an end, to greater security in old age or whatever... those people need a different level of protection from the people who [invest etc] as a business": evidence to the Joint Committee, *op cit* n 36 (vol. II), para 141.

⁴⁶ See Income and Corporation Taxes Act 1988, s 634(6).

These singular factors carry particular ramifications both for the objectives of regulation and the design of regulatory institutions. It has been argued that “the most powerful economic justification for public intervention is the correction of market failure”.⁴⁷ The goal of regulation here should be to reduce the prospect of such failure, by focusing closely upon controlling the risks to which consumers remain exposed as a result of the location of the product in the financial marketplace. This will maximise the chance that the PP will achieve its social policy aims. It is not enough simply to prevent the consumer from “being made a fool of”.⁴⁸ The latter objective may be appropriate for other financial products, but the individual, social and political consequences of failure of the PP are such that only very low levels of risk are acceptable. The regulator must accordingly attempt to position the PP purchaser in a low-risk environment, which will entail assuming a relatively unsophisticated customer with a low tolerance of risk and a risk-averse investment portfolio.⁴⁹

The need for this form of regulation distinguishes the PP from other financial service products.⁵⁰ Indeed, it can be argued that the use of PPs as an instrument of government policy, to meet a particular and pervasive social risk, gives rise to ‘public interest’ considerations more closely analogous to those that exist in utility products:

“if people are to have a reasonably good quality of life, gas and water services must be provided. I simply contend that some financial service products – such as those producing a decent retirement income – are also necessary in ensuring that people have a good quality of life”.⁵¹

This is especially so given their continued role in supplementing state provision in future government plans for the welfare system.⁵² All of these factors point to the need for a regulator to adopt an approach which *differentiates* between the PP and other financial service products, recognising its unique characteristics and purpose and managing risks accordingly.

It has been argued that the regulatory institutions governed by the Financial Services Act 1986 “did not have a detailed awareness of the particular nature of PPs” and hence did not conceive the PP product as problematic.⁵³ The necessary level of differentiation between products was not achieved, with the regulators monitoring “the whole of a firm’s business across the spectrum of its regulated activities”⁵⁴, rather than recognising the distinctive qualities and

⁴⁷ Bishop *et al*, *The Regulatory Challenge* (1995), p 3.

⁴⁸ See Gower, above n 26.

⁴⁹ See further Miles and Timmermann, *op cit* n 10 at 277, who argue for “some role for government to play in setting up guidelines for diversification and risk management service”.

⁵⁰ It is acknowledged that there are other financial products that aim at retirement provision; it is submitted that these should be subjected to the same risk analysis and regulatory regime suggested here for pensions.

⁵¹ Love, 334 HC Deb 97 (28 June 1999).

⁵² See below, n 118. Also note that it is difficult for IFAs to advise moving out of PPs, for example because of the controls on ‘churning’: see SIB core rule 26(1) – hence advice is likely to be to retain the product once it is purchased.

⁵³ Black and Nobles, *op cit* n 6 at 813.

⁵⁴ *Ibid*.

difficulties associated with PPs. This can be seen as a consequence of the institutional structure of regulation, which entailed the PP product being overseen by two SROs, both responsible for a variety of other financial service products in which higher levels of risk were acceptable.⁵⁵ A further difficulty was the absence of a statutory principle of consumer protection in the 1986 Act⁵⁶, which might have shifted the regulatory focus away from the *investor* towards the more vulnerable *consumer*, prepared to tolerate only low levels of risk.

It may be over simplistic to ascribe the mis-selling of personal pensions solely to the inappropriate nature of the regulatory style and structure emerging from the 1986 Act⁵⁷; nor do we take issue with government regulation *per se*. It is our contention, however, that the previous regime was seriously flawed in its failure to account for and react to the distinctive characteristics and specific risks inherent in the PP product. The question which must now be addressed is: will the new regulator, the Financial Services Authority (FSA), be any better placed to do so?

THE FSA: THE RIGHT REGULATORY TOOL FOR THE JOB?

1. Creating the FSA

The radical overhaul of the structure of financial services regulation which was announced by the new Chancellor in May 1997 took commentators by surprise.⁵⁸ Arguing that “the regulatory structure introduced by the Financial Services Act 1986 is not delivering the standard of supervision and investor protection that the industry and the public have a right to expect” and criticising the “inefficient” division of responsibility between SIB and the other regulatory bodies, he announced that work was to begin on legislation to simplify the regulatory system.⁵⁹ The new regulator was to take the form of a statutory ‘super-SIB’, with responsibility for all aspects of banking and financial services regulation, although the Bank of England would retain responsibility for overall systemic regulation.

The single regulator, christened the Financial Services Authority, came into being on 28 October 1997. The regulatory functions and powers previously exercised by the institutions governed by the 1986 Act will be vested in the FSA by means of a Financial Services and Markets Bill, likely to receive

⁵⁵ Although the replacement of Fimbria and Lautro by one regulator, the Personal Investment Authority (PIA), in 1994 may have reduced duplication of tasks, the responsibilities of this agency continued to extend well beyond the PP context.

⁵⁶ An omission particularly criticised by the National Consumer Council: see *Financial Services and Markets Bill: Response to the Treasury’s consultation on the draft Bill* (1999), p 3.

⁵⁷ See Davies, in evidence to Treasury Select Committee, Third Report, *Financial Services Regulation* HC 73-II (1998-9), para 216; and for a discussion of other contributory factors, see Black and Nobles, *op cit* n 6. It is important to note, however, the use of the mis-selling episode as a key argument for regulatory reform: see below, n 66 and accompanying text.

⁵⁸ See *eg* “Labour turns to the City”, *The Economist* (24 May 1997), p 15.

⁵⁹ Brown, 294 HC Deb 509 (20 May 1997).

Royal Assent in spring 2000.⁶⁰ The legislation, and hence the regulatory regime, extends to Northern Ireland.⁶¹ The Authority operates in effect as a statutory arm's-length government agency with a governing Board appointed by the Treasury and accountable to Parliament; but in legal form it is a private company limited by guarantee, funded by levies on the financial services industry.

2. The FSA and PPs: grounds for optimism

An evaluation of the likely effectiveness of the new regulator in addressing the particular challenges presented by the PP product necessitates discussion of the reasons for its creation, its statutory objectives and the approach that it proposes to adopt in the exercise of its powers.

The Government identified two broad rationales for overhauling the regulatory regime, and for the model adopted. Treasury ministers argued that a single regulator would reduce duplication of effort (and hence compliance costs) and better reflect the structure of a modern, global and integrated financial services industry, in which distinctions between types of institution are increasingly blurred.⁶² Regulatory reform was therefore linked to prosperity and international competitiveness, as well as Labour's broader 'modernising' agenda.⁶³ This justification, which might be described as emanating from *structural change*, thus related to wider developments in financial markets, rather than the PP 'problem' in particular.⁶⁴

The second – and here more instructive – rationale was *systemic confidence*, based on a critique of the previous 'self-regulatory' regime. In support of the argument that self-regulation, while attempting to protect the interests of the financial services industry as well as those of the public, ultimately did neither,⁶⁵ the Government marshalled as evidence incidents in which past regulatory performance appeared defective, chief among which was mis-selling:

⁶⁰ HC Bill 121 (1998-9). It received its First Reading on 17 June 1999, by which time it had already been subject to extensive consultation and pre-legislative scrutiny: see Joint Committee, *op cit* n 36, para 2. This process may in part have been designed to answer criticism that the effective creation of the FSA before the passage of its enabling legislation precluded adequate debate on the reforms: see Taylor, "Unanswered questions for the UK's new leviathan" (1997) 2(4) *The Financial Regulator* 42.

⁶¹ Clause 364. Excluded are provisions regarding open-ended investment companies, which are not relevant here.

⁶² See Brown, above n 59 at 510; Milburn, 334 HC Deb 37 (28 June 1999).

⁶³ See especially Byers, speech to Corporation of London Finance Committee, Treasury Press Release 201/98 (26 November 1998).

⁶⁴ Although implications exist for the speed and effectiveness of the regulatory response to issues such as mis-selling: see below, n 86 and accompanying text. For further discussion, see Goodhart *et al*, *Financial Regulation: why, how and where now?* (1998), ch.8.

⁶⁵ See O'Brien, "Regulatory change – revolution or reform?" (1996) 1(3) *The Financial Regulator* 17 at 18; Darling, speech to FSA European Conference, Treasury Press Release 201/98 (26 November 1998), para 30.

“the scale of personal pensions mis-selling has made it absolutely critical that, for consumer confidence, there now has to be a more transparent, a much simpler system of regulation that is not self-regulation. Personal pensions mis-selling brought the whole ethos of self-regulation into disrepute”.⁶⁶

Accordingly, the professed goal of reform was to “increase public confidence in the regulatory regime”⁶⁷ and in the markets, by restructuring regulation so that it was perceived as more independent, transparent, accessible and robust in its responses to market failure.⁶⁸ These qualities provided a link with the structural objective of simplification and pointed towards the creation of a statutory single regulator which could react more efficiently and effectively to crises such as pensions mis-selling.

The issue of public and consumer confidence lies, therefore, at the very heart of the FSA’s *raison d’être*, and the mis-selling episode offers both a specific justification for dismantling the previous regulatory regime and an illustration of the sort of incident that must be avoided in future. Consequently, it might be expected that the FSA will focus on measures to enhance public confidence and to protect the consumer, with those arising in the pension context particularly highlighted.

There is strong support for the view that that “the FSA’s primary remit will be to protect the consumer”.⁶⁹ The Financial Services and Markets Bill specifies the promotion of “public understanding of the financial system”⁷⁰ and the “appropriate degree of protection of consumers”⁷¹ as key objectives of the new regulator,⁷² while the FSA has described consumer protection as “at the heart of the work of the Authority”⁷³ and as the “red meat of our task”⁷⁴. This might lead one to predict that the FSA will pay close attention to the effective management and reduction of the risks faced by consumers within the financial marketplace – for example, by redressing asymmetries of information.

⁶⁶ Liddell, in evidence to Treasury Select Committee, *op cit* n 5 (vol. II), para 21; see also Brown, letter to Chairman of SIB, Treasury Press Release 49/97 (20 May 1997): “The government is committed to reform... to reduce the chance of events such as the mis-selling of personal pensions happening again”; also Milburn, above n 62 at 38; and for comment, Taylor, *op cit* n 60 at 43, 45.

⁶⁷ Brown, above n 59.

⁶⁸ See Foreword (Brown), *Financial Services and Markets Bill: A Consultation Document (Part 1)* (1998).

⁶⁹ Taylor, *op cit* n 60 at 43. Such a focus is perhaps unsurprising, given the shift in regulatory attention towards consumer and public interest issues in recent years, reflected elsewhere by the priority given to consumer interests in proposals for the reform of utilities regulation: see DTI, *A Fair Deal for Consumers* (1998; Cm 3898).

⁷⁰ Clause 4.

⁷¹ Clause 5(1).

⁷² Additionally, clause 9 of the Bill requires the establishment and maintenance of a panel “to represent the interests of consumers”.

⁷³ FSA, *op cit* n 25, para 43

⁷⁴ Davies, speech at FSA Launch Conference (28 October 1997).

But by itself, a commitment to ‘protecting the consumer’ – although preferable on our analysis to an approach that simply seeks to ‘prevent him from being made a fool of’ – may not be sufficient. We have argued above that the nature and purpose of the PP product and the consequent need for the reduction of risk, distinguish it from other transactions within the financial services market. It is important, therefore, that the FSA should take a *differentiated* approach to consumer protection, allowing it to intervene to a degree which would be inappropriate with other products in order to minimise the risks that might prevent the PP from fulfilling its objective. To what extent can the Authority be expected to adopt the necessary varied and flexible approach to its regulatory tasks?

A definitive response cannot be given at this stage, but a number of early indications exist. The Bill requires the FSA to have regard to the principle of *caveat emptor*⁷⁵, seen as necessary to avoid the risk of “moral hazard”⁷⁶; but in considering the degree of consumer protection which is appropriate, it will attempt “to ensure that they [consumers] are not exposed to risks that they should not reasonably be expected to assume”.⁷⁷ This is reflected in provisions requiring the Authority to have regard to differing degrees of “risk involved in different kinds of investment or other transaction” and of “experience and expertise which different consumers may have in relation to different kinds of regulated activity”, together with consumer needs for advice and accurate information.⁷⁸ Here, therefore, is a statutory manifestation of the view that the variety of products and consumers demands “an appropriate level of consumer protection that will be greater in some cases than others”⁷⁹, leaving it open to the FSA to apply differing levels of regulatory protection when undertaking its statutory functions. Thus, for example, the marketing and sale of life assurance could be regulated in a different fashion to that of PPs.

The FSA has already indicated that it is “committed to a flexible and differentiated approach”,⁸⁰ and has shown apparent awareness of the particular importance of such an approach in the PP context:

⁷⁵ Clause 5 (2) (d).

⁷⁶ *Ie* not requiring consumers to accept responsibility for actions would “give them no incentive to take appropriate decisions”: Treasury, *Financial Services and Markets Bill – Progress Report* (1999), para 4.11.

⁷⁷ FSA, *Financial Services Authority: an outline* (1997), para 2.

⁷⁸ Clauses 5 (2) (a), (b) and (c).

⁷⁹ Treasury, *op cit* n 76. Indeed, it is significant that the Government rejected calls from bodies such as the National Consumer Council (*op cit* n 56, p 3) to distinguish between ‘retail’ and ‘wholesale’ sectors, viewing this as “too blunt an instrument” (Treasury, *ibid*, para 4.8), which failed to take account of “a spectrum of types of customer within these broad categories”: *Financial Services and Markets Bill – Government Response to the Joint Committee’s First Report* (1999), para 2. See above, n 36 and accompanying text.

⁸⁰ FSA, *Designing the FSA Handbook of Rules and Guidance* (1998), p 6; also *Differentiated Regulatory Approaches – Future Regulation of Inter-professional Business* (1998), pp 5,7. Note also its proposals to utilise ‘risk-based’ supervision, focusing on the nature of the activity, the extent of risk within firms and markets, the quality of management controls and the sophistication of the consumers involved: *op cit* n 77, para 3.

“the personal pension is at one extreme of the spectrum of financial products. There are other savings and investments, such as simple deposit accounts which are much easier to understand. . . . So it is clear that the nature of any regulation of the retail market should vary considerably across products, and in some cases very little regulation is needed”.⁸¹

This would suggest, therefore, that we can expect the new regulator to exercise its statutory powers and duties in an adaptable manner, varying its style and level of regulation to respond adequately to the differing levels of risk existing within the financial services sector, and providing a greater level of protection in cases where – as with PPs – only very low levels of risk are acceptable. For the reasons previously advanced, such an outcome would be welcome.

3. The FSA and PPs: potential pitfalls

This outcome is, however, far from guaranteed. Despite the statutory provisions and early regulatory assurances, concerns have been voiced that the “FSA will fail to differentiate adequately. . . in the way that it approaches supervision and regulation”, especially given the broad definition of ‘consumer’ contained in the Bill.⁸² If this proves to be so, there is a danger that the particular characteristics and problems associated with the PP product and consumer may once again be left unrecognised. We turn, therefore, to examine the factors lending credence to this gloomier prognosis.

The ‘structural change’ rationale presented the new regulatory regime as more efficient and a better reflection of developments in the financial services industry.⁸³ Here, in arguing that “creating a single regulator with a single authorisation process, a single compensation scheme, a single ombudsman, and a single appeals tribunal, [will] reduce the amount of regulation while at the same time providing for greater accountability”,⁸⁴ the Government offered an *integrative* justification for regulatory reform. This was further evident in the emphasis placed upon the need to avoid the regulatory “turf wars”⁸⁵ that supposedly characterised the previous regime. A single regulator was regarded as being in a position to respond more effectively and speedily to incidents such as mis-selling:

“There can be little doubt that the regulatory response to previous mis-selling would have been more coherent, and more rapid, had it been managed in a single centre rather than being the responsibility of a number of separate organisations. Individuals in the regulatory network tried very hard to achieve coherence and co-operate with each other; the framework within which they were operating did not reward those efforts. The establishment of

⁸¹ Davies, Henry Thornton Lecture, FSA Press Release 87/98 (4 November 1998).

⁸² Joint Committee, *op cit* n 36, para 29. See further below, n 94.

⁸³ Above, n 62 and accompanying text.

⁸⁴ Byers, above n 63.

⁸⁵ See Alexander, “A view from the market” (1997) 2(3) *The Financial Regulator* 28 at 28-9.

the FSA as the single statutory regulator is putting an end to this regulatory fragmentation”.⁸⁶

Pressure for an integrated form of regulation, reflected in the rhetoric of the “one-stop shop”,⁸⁷ has therefore been imprinted strongly within the FSA’s collective genes from birth, and represents a strong countervailing trend to the differentiated approach discussed above.

Moreover, it is likely that as it grows into maturity, the Authority will demonstrate ever more powerful integrative tendencies. Organisational theory teaches that, in order to contain fragmentation, successful large organisations need to adopt integrative managerial devices to co-ordinate the range of tasks.⁸⁸ When this factor is viewed alongside the Authority’s deployment of a ‘lead supervisor’ approach to the regulation of financial conglomerates,⁸⁹ and the impetus provided by regulated firms preferring a ‘level playing field’ to a differentiated style which might confer a regulatory advantage on their rivals,⁹⁰ it would appear that there is a strong possibility of the development of a monolithic organisational ‘culture’ or ‘ethos’ within the FSA:

“Large organisations tend to develop a corporate approach; indeed they usually take pride in doing so. A real danger is that the FSA may try to deal in much the same way with very different market sectors, and with different customers with different requirements”.⁹¹

There are reasons to believe, therefore, that FSA regulation may not differentiate to the extent presently asserted, and that a more homogeneous approach may steadily evolve. The issue here is the nature of any ‘culture’ that emerges and its likely effect upon the regulator’s handling of the PP product. On one level, it might be argued that an undifferentiated regulatory approach need not in itself be unduly troubling, provided it is one that remains oriented towards protecting consumers from risks faced in the financial markets.⁹² Given the prominence accorded to consumer protection

⁸⁶ FSA, memorandum of evidence to Treasury Select Committee, *op cit* n 66, para 25; see also Davies, *op cit* n 57, para 219. The FSA has taken steps towards such a co-ordinated response by establishing, with PIA, a single integrated pensions review division: see FSA Press Release 07/97 (9 December 1997).

⁸⁷ See Davies, in evidence to Treasury Select Committee, *ibid*, para 195; Milburn, above n 62 at 35.

⁸⁸ See Handy, *Understanding Organizations* (4th ed., 1993), pp 203-9.

⁸⁹ This entails an individual or team co-ordinating the activities of other FSA employees working within the agency’s different divisions, and taking responsibility for the consolidation of the supervisory relationship between such a firm and the Authority: see FSA, *op cit* n 25, para 35.

⁹⁰ See Alexander, *op cit* n 85 at 29: “we [NatWest Group] obviously want the FSA to develop a common culture and style so as to be consistent in its approach”.

⁹¹ Lloyd, 334 HC Deb 80 (28 June 1999). The FSA has also emphasised the “benefits of a single regulatory culture”: see *The Financial Services Authority: Information Guide* (1999), p 7.

⁹² The discourse of ‘consumer’ – as distinct from investor – which underpins the Bill would seem to focus regulatory attention on market risks faced by individuals: see Davies, above n 74.

issues in the creation of the FSA, this would indeed seem the most likely short-term outcome.

However, as argued above, a ‘consumer orientation’ is not enough in itself: if the particular risks inherent in the PP are to be effectively managed so that its social policy objective is realised, it is important that the regulator continues to differentiate between the PP purchaser and other types of consumer: in short, that it recognises the existence of a ‘spectrum’.⁹³ The integrative pressures previously outlined, coupled with the absence of a statutory distinction between types of ‘consumer’, create anxiety that the FSA may move towards a uniform approach to its consumer protection responsibilities across markets and products.⁹⁴ In view of the retention of the *caveat emptor* principle, any such approach is most likely to be slanted towards an individual who is prepared to accept moderate levels of risk and responsibility, rather than the PP purchaser located in a ‘low-risk environment’: in other words, that “we will finish up with a regime fixed in the middle”.⁹⁵

Furthermore, concern has been expressed that, over a longer period, there is a danger of ‘regulatory capture’ by the financial services industry.⁹⁶ This is especially so given the difficulty of organising effective representation of consumers due to the disparate nature of their interests (and the benefits gained) and the lack of resources comparable to those of producers.⁹⁷ The Authority may consequently develop an industry-oriented ethos, shifting its activity away from areas – such as PPs – where consumer protection issues are paramount, as well as exhibiting greater tolerance toward practices that create risks for consumers. Such an outcome may be all the more likely once the mis-selling episode has faded from media and political memories.

4. The FSA and PPs: conclusion

It is apparent from this discussion that it is, at present, impossible to be certain of the manner in which the FSA will approach its tasks. Initial indications point toward a differentiated style,⁹⁸ allowing room for cautious optimism that there will be proper regulatory recognition of the distinctive

⁹³ See above, n 79.

⁹⁴ See Treasury, *op cit* n 76, para 4.8; National Consumer Council, *op cit*, n 56, Hall (National Consumer Council) in evidence to Joint Committee, *op cit* n 45, para 141.

⁹⁵ Cable, HC Standing Committee A (15 July 1999).

⁹⁶ See Consumers’ Association, memorandum of evidence to Treasury Select Committee, *op cit* n 66, para 4: “It is by no means certain that the FSA will deliver an effective regulatory regime according to consumer principles. There is a real fear that the consumer interest will be marginalised by the industry lobby in such a huge organisation”; National Consumer Council, *op cit* n 56, pp 10-11: “over time, the financial services industry. . . may gain excessive influence over the Authority”. Producer dominance has been viewed as a key feature of ‘self-regulation’: see Alcock, “The New Financial Services Authority and Consumer Protection” (1998) 2 *Company Financial and Insolvency Law Review* 88 at 91-2.

⁹⁷ See especially Olson, *The Logic of Collective Action* (1965), pp 165-6, endorsed in the context of the Bill by McKechnie (Consumers’ Association), in evidence to Treasury Select Committee, *op cit* n 57, para 11.

⁹⁸ See above, n 80 and accompanying text.

problems and characteristics of the PP. However, a powerful counter-trend towards integration is also evident from the FSA's origins, which may well grow stronger as the Authority evolves. If this tendency proves dominant, regulation of the PP product may once again become subsumed within wider financial service matters, with the likely consequence that the specialised risk management needed to achieve its distinctive policy goal is neglected.

That the FSA is conscious of these issues is apparent from its identification of the importance of "be[ing] alert to the need to strike a balance between the drive for integration and the maintenance of flexibility and appropriate differentiation".⁹⁹ Nonetheless, given the concerns expressed here, it seems appropriate to put forward possible alternatives to the new regulatory regime, which might be better suited to the particular challenges offered by the PP.

TOWARDS A NEW REGULATORY STRUCTURE?

There are clear arguments for the retention of an arm's-length, statutory agency, whose perceived distance from industry interests is likely to be a key factor in reinstating public confidence in the personal pensions market.¹⁰⁰ However, as previously discussed, "there are greater pressures towards undue uniformity within a single regulator – where every difference needs justification – than in specialist regulators which are clearly freer to develop their own distinctive approaches".¹⁰¹ Given the importance of differentiating PPs from other financial products, we therefore believe that the model should be such as to allow for greater regulatory specialisation.

Comparison may be drawn with the utilities sector, where bodies such as OFTEL and OFWAT regulate within a statutory framework set up for each individual industry. The objectives of a 'watchdog' in the competitive market for PPs might appear to be somewhat different from those of the utility regulators, whose major goals have been to stimulate competition and prevent abuse of monopoly power by the former nationalised companies; but there are indications that the 'utility watchdog' analogy has gained some ground.¹⁰² However, an even closer parallel exists within the pensions sector itself where the Occupational Pensions Regulatory Authority (OPRA) focuses solely on employer-run schemes, under the provisions of the Pensions Act 1995. An argument might be made for extending OPRA's regulatory responsibilities to PPs: this would detach the regulation of the "very particular business"¹⁰³ of pension selling from that of other financial products, while simultaneously emphasising the comparable social policy goals of the differing pension vehicles. However, care would need to be exercised as the industrial relations and trust-based form of the OP raises quite distinct legal and other problems from the PP, which centres on arms-

⁹⁹ FSA, *op cit* n 25, para 67.

¹⁰⁰ See Hedges, *Pensions and Retirement Planning*, DSS Research Report 83 (1998).

¹⁰¹ Lomnicka, "Reforming UK Financial Services Regulation: The creation of a Single Regulator" [1999] *JBL* 480 at 489.

¹⁰² See Love, above n 51 and accompanying text; also the Treasury Select Committee's suggestion that promotion of competition might become a statutory objective of the FSA: *op cit* n 57 (vol. I), para 23, subsequently rejected by the Government: Treasury, *op cit* n 76, para 4.16.

¹⁰³ Black and Nobles, *op cit* n 6 at 791.

length contract – hence different skills and regulatory approaches may well be needed.¹⁰⁴

The application of some form of sector-specific regulatory model to the PP situation would have a number of advantages. Most significantly, it would facilitate the development of a level of specialised agency expertise. This is perceived as being a key strength of the utility regulators,¹⁰⁵ whose importance in the PP context has been emphasised by Black and Nobles. They claim that a feature of mis-selling was “regulatory blindness. . . No-one looked at pensions as posing particular problems because no-one knew or thought to look”, and conclude that “the regulator needs to have specialist knowledge of individual product and business areas”.¹⁰⁶ The ‘technical’ complexity of the PP product, especially as regards issues such as the tax regime, buying out, obtaining transfer values, the fee and commission structures combined with perhaps unfamiliar investment media and the inevitably long-term and retirement-focused nature of the PP – in addition to the distinctive social policy issues discussed above – presents a particularly strong case for specialisation here.

Such expertise may, of course, be attainable within the much larger structure of the FSA, but it is submitted that a smaller, focused agency would be better placed to identify the particular risks associated with PPs, to devise an appropriate response to instances of market failure and to offer an environment in which the special knowledge and skills needed to regulate this product can be developed. As argued in parliamentary debate, “the problem with one-stop shopping is that the shop assistants may not know well the needs of their different types of customer or the characteristics of their different suppliers”.¹⁰⁷

Furthermore (and contrary to the argument for regulatory simplification advanced above),¹⁰⁸ it may also be that a specialist agency can achieve a swifter and more flexible response to problems in the pensions arena than a larger, more bureaucratic organisation.¹⁰⁹ And, while regulatory capture remains a theoretical possibility for such a body, just as with the FSA, the experiences of the utility regulators and OPRA to date suggest that such an outcome is unlikely.¹¹⁰

Two objections to such a reform of the structure of regulation may be raised, and met. Firstly, it would represent a return to a more complex regime, with

¹⁰⁴ Black and Nobles, *ibid* at 793, identify the failure to compare the benefits of OPs with those of PPs as an illustration of mis-selling, but also stress the extent of the differences and the consequent need to develop specialised knowledge: at 790, 794-5, 817-8. It is notable that the Government has decided against absorbing OPRA within the FSA and thus subjecting the two pension vehicles to the same regulatory regime: see Joint Committee, *op cit* n 36, Appendix 1.

¹⁰⁵ See Harlow and Rawlings, *Law and Administration* (2nd ed., 1997), p 310.

¹⁰⁶ Black and Nobles, *op cit* n 6 at 789-90.

¹⁰⁷ Lloyd, above n 91 at 79-80.

¹⁰⁸ See above, n 86 and accompanying text.

¹⁰⁹ For discussion of the view that the FSA’s size is likely to result in unresponsiveness, see Davies, above n 74.

¹¹⁰ On the utility regulators, see Corry *et al*, *Regulating Our Utilities* (1994), pp 74-5.

firms regulated by different agencies according to whether they are conducting pension or other types of financial business. This apparently offends against one of the Government's main rationales for structural change, reflected in the "one-stop shop" rhetoric. There are already, however, some significant exclusions from the FSA's regulatory scope.¹¹¹ Moreover, it is interesting that recent proposals for the creation of 'stakeholder' pension schemes envisage regulation by two bodies: the FSA will regulate the marketing of schemes and provision of advice, while OPRA will oversee the operation of schemes.¹¹² These factors suggest that the Government does not regard the 'single regulator' solution as an article of faith, and that it is prepared to contemplate variations on this structure where appropriate.¹¹³ Our view is that PPs *are* sufficiently distinctive to require just such a departure from the regulatory norm.

Secondly, one might query the practicability of setting up such a specialised agency, given that the FSA is already discharging its regulatory functions within the PP market. Adjusting the regulatory structure once more would delay the legislative process and cause considerable disruption for firms, consumers and regulatory staff alike.¹¹⁴ However, acceptance of this argument may result in "the organisational tail. . . wagging the legislative dog":¹¹⁵ it is surely preferable to address possible deficiencies at a relatively early stage rather than to revisit them in subsequent legislation. In any event, it would be possible to allow for creation of a specialised regulator without significantly delaying the passage of the Financial Services and Markets Bill by the simple expedient of inserting an amendment allowing the FSA to delegate its powers and duties to another body which met prescribed statutory standards.¹¹⁶ This would allow for dispassionate, carefully-considered theorising focusing upon the type of regulator which best suits pension provision in general and the PP market and the regulatory difficulties it presents in particular.¹¹⁷

¹¹¹ Notably mortgages, which have not previously been subject to the statutory financial services regime: see Joint Committee, *op cit* n 36, paras 63-89. This exclusion is presently under review: see Treasury, *Regulation of Mortgages: A Discussion Document* (1999).

¹¹² See DSS, *Stakeholder Pensions: Regulation, Advice and Information – The Government's Proposals* (1999); Welfare Reform and Pensions Bill, HC Bill 44 (1998-9), clause 2.

¹¹³ See DSS, *ibid*, paras 10-11, where arguments for a single regulator are explicitly rejected.

¹¹⁴ See Treasury Select Committee, *op cit* n 102, para 53.

¹¹⁵ Taylor, in evidence to Treasury Select Committee, *op cit* n 57, para 184.

¹¹⁶ Taylor, memorandum of evidence to Treasury Select Committee, *ibid*, para 7.

¹¹⁷ The absence of such theorising in the approach to regulatory design is not unusual in the UK. It has been observed that "it is characteristic of our administrative law that machinery should evolve in piecemeal fashion and with no single set of clearly defined objectives. Utilities regulation has been peculiarly susceptible, given the scale and complexity of the privatisation process, as also a steep learning curve for government and agencies alike. Diversity in powers and performance is one trait which the regulation shares": Harlow and Rawlings, *op cit* n 105, p 318.

CONCLUSION

The argument advanced in this article is that the *social* goal of pension provision – to ensure that as many as possible attain an adequate level of income on retirement – gives rise to distinctive challenges for regulation of the PP and creates a particular need for effective risk management to ensure that the objective is fulfilled. In spite of its initial assurances, there are reasons to question whether, especially in the long term, FSA regulation will prove to be sufficiently sensitive to these issues. Such anxiety points us toward the creation of a focused regulatory agency in which the skills necessary to manage this distinctive and complex product can be evolved and which avoids the hazards of an undifferentiated approach.

Of course, institutional reform will not in itself provide a panacea to the problems that have beset the personal pension market. It is important that the creation of an appropriate regulatory structure is combined with the promulgation of rules which properly address the potential abuses which may occur, coupled with a preparedness to enforce these rules rigorously, imposing sanctions where necessary. The Labour Government, mindful of the continued role for the PP in its welfare policy,¹¹⁸ has adopted a number of additional measures in an attempt to “put right the wrongs of pensions mis-selling”.¹¹⁹ An equally robust approach will be required of the FSA, or any alternative regulator, if similar incidents are to be avoided in the future.

¹¹⁸ See DSS, *op cit* n 1, ch.3, para 9: “appropriately regulated, personal pensions will be the right option for some, mainly higher paid people”. Note also the Government projection that the present 60:40 ratio of state to private provision will be reversed over the next fifty years: *ibid*, ch.12, para 4.

¹¹⁹ See Liddell, 301 HC Deb 155-7 (18 November 1997). Note also various FSA advertising campaigns: see FSA Press Release 18/98 (14 April 1998), 001/99 (5 January 1999).

PAST HUMAN RIGHTS VIOLATIONS: TRUTH COMMISSIONS AND AMNESTIES OR PROSECUTIONS

Mr Justice Richard J. Goldstone, Constitutional Court of South Africa

*The MacDermott Lecture 1999 **

It is a great honour to have been invited to deliver this 1999 Annual MacDermott Lecture. It is a particular privilege because of the eminence of Lord MacDermott, not only in this country, but internationally. And it is a particular pleasure because it has brought my wife and me to Belfast for the first time, and it has enabled us to meet Lord Justice MacDermott and Lady MacDermott and members of the MacDermott family.

I was going to begin by saying that I was not coming here to suggest that you could learn something from South Africa. But I decided not to because of the memory it immediately evoked of the many people who came from abroad to South Africa during our dark days of apartheid and assured us that they were not coming to tell us that we could learn from their experience. Of course they meant the very opposite. And I am sure many people come here, and have been doing so for many years, telling you that they are not here to tell you anything that you should learn about their condition or about their solutions. But I do think we can all learn some things from each other. Certainly I am happy and proud to say as a South African that, countries that have the sorts of difficulties we experienced, and unfortunately, there are too many of those countries, can at least learn that solutions can be found even for situations that appeared as intractable as bringing apartheid to a relatively peaceful end.

The broader problem, contained in the title of this lecture, that is, what is to be done about past human rights violations, has become fairly common. We should rejoice in this as it is a problem which presents itself as countries move out of oppression, whether they are communist societies, military dictatorships in Latin America or apartheid in South Africa. These are countries which have moved from oppression to some form of democracy, and the question facing those societies is what they should do about past criminality, past human rights abuses.¹ Generally the countries concerned

* My thanks to Ms Imelda McAuley of the School of Law, Queen's University, Belfast for providing footnote references to accompany the text of this lecture.

¹ On this question, see for example, N Roht-Arriaza, "State Responsibility to Investigate and Prosecute Grave Human Rights Violations in International Law", 78 *Calif LR* (1990) 449; D. Orentlicher, "Settling Accounts: The Duty to Prosecute Human Rights Violations of a Prior Regime", 100 *Yale LJ* (1991) 2537; J. Benomar, "Confronting the Past: Justice After Transitions", 4 *Jnl of Democracy* (1993) 3; N. Kritz (ed), *Transitional Justice: How Emerging Democracies Reckon with Former Regimes* (1995, United States Institute of Peace, Washington D C); M. Scharf, "The Letter of the Law: The Scope of the International Legal Obligation to Prosecute Human Rights Crimes", 59 *L & Contemp Probs* (1996) 41; J. Mendez, "Accountability for Past Abuses", 19 *Hum Rts Q* (1997) 255; Osiel, "Why Prosecute? Critics of Punishment for Mass Atrocity", 22 *Hum Rts Q* (2000) 118.

have been faced with three broad choices. First, they can forget about the past and enter into a period of national amnesia. Secondly, they can systematically prosecute perpetrators of criminality. Thirdly, they can establish a truth commission process which is really a compromise between doing nothing on the one hand and prosecuting on the other. Now, the choice of doing nothing has appealed to some countries. It was very appealing to many in South Africa and particularly for the former leaders of the apartheid government and leaders of the security forces. And of course they made a very beguiling case. They said: “we have got so much to do for our future, we have so much to make up for, we have to redistribute wealth, we have to get rid of all of the evils of the apartheid system, why waste our time looking over our shoulders to the past. Let us forget about that. We are turning over a new leaf and let us get on with building a happier future for all of our people.” One thing I have learnt in South Africa, and again when I was investigating war crimes in Bosnia and the former Yugoslavia, and also in Rwanda, is that countries choosing the amnesia route have condemned their societies to more violence and more unhappiness – sooner rather than later. In the former Yugoslavia, the history lessons that I received on every visit to Belgrade, Zagreb and Sarajevo taught me what happens to a society that does not bring justice, some form of justice, and some form of acknowledgement, to the victims.

I read recently that the former Yugoslavia has more history than their people can consume. I thought that was certainly a neat way of encapsulating what I experienced. At the first meeting I held in Belgrade, I had scheduled a half hour meeting with the Minister of Justice and the Foreign Minister. The first forty-five minutes of the thirty minute-meeting were consumed with a long history lesson about the terrible things that had been suffered by the Serb people at the hands mainly of the Croats, but also at the hands of the Muslims. And the history lesson began with the battle of Kosovo in 1389 and ended with the Second World War and the loss of life of hundreds of thousands of Serbs at the hands of the Croatian Ustashe. And I went to Zagreb and I got a similar history lesson about the terrible things that the Serbs and the Muslims had done to the Croats. And in Sarajevo I had the story from the Muslim side. The histories never dovetailed, they never even intersected, because each of these societies was completely embroiled in its own history. And the terrible things that we are reading right at this moment about Kosovo might well have been averted had that history not been written the way it was, but instead there had been some attempt to uncover what really happened over the centuries, to uncover a sense of shared history.

So the objective really, if one is sensible, is to bring some form of justice to societies that have endured this sort of history of human rights violation. Prosecutions are obviously the preferred solution. In any ideal society, in any decent society, in any normal society, victims are entitled to full justice and full justice means prosecution and punishment. But, unfortunately, one is dealing with such massive numbers. In South Africa there were tens of thousands of victims of serious crime, committed in the name of apartheid. In Rwanda there were hundreds of thousands of victims and hundreds of thousands of perpetrators, people who organised the genocide and terrible

crimes against humanity.² In the former Yugoslavia similarly, and in Bosnia alone, hundreds of thousands of people were forced to become refugees, many tens of thousands of women were raped and even more tens of thousands of people – men, women and children – were tortured, murdered and ethnically cleansed. In these situations no criminal justice system in any country can cope with prosecuting the criminals. You could opt for Nuremberg Trial-type prosecutions and place the most important leaders on trial and have them serve as an example to the victims. In some situations, and certainly after the Second World War, that was a rational and sensible solution. Yet the Nuremberg Trials have often been criticised as an impure form of justice, a “victor’s justice”. But one must bear in mind when looking at Nuremberg that the choice facing the victorious allied powers was one between lining up the Nazi leaders and executing them, military style, which is what Stalin wanted and initially what Winston Churchill wanted. Fortunately, good sense prevailed and the victors decided that there would be at least a form of justice, a decent trial at which the guilt of the accused would have to be proved beyond a reasonable doubt.

The Truth Commission route has really emerged as a political necessity more than the ideal solution.³ The Truth Commission which most influenced South Africa was the Chilean model.⁴ Chilean President Aylwin, as the rule of General Pinochet was ending, promised the Chilean people that there would be a Truth Commission, that the past would not be covered up, that the disappearances, at least, of so many thousands of Chileans, would be

² On the debate over whether a truth commission should be established in Rwanda, see J. Sarkin, “The Necessity and Challenges of Establishing a Truth and Reconciliation Commission in Rwanda”, 21 *Hum Rts Q* (1999) 767.

³ On truth commissions generally, see P. Hayner, “Fifteen Truth Commissions - 1974 to 1994: A Comparative Study”, 16 *Hum Rts Q* (1994) 597; J. M. Pasqualucci, “The Whole Truth and Nothing But The Truth: Truth Commissions, Impunity and the Inter-American Human Rights System”, 12 *Boston Univ Int L J* (1994) 321; T. Buergenthal, “The United Nations Truth Commission for El Salvador”, 27 *Vanderbilt J Transnat'l L* (1994) 497; M. Popkin and N. Roht-Arriaza, “Truth as Justice: Investigatory Commissions in Latin America”, *L & Soc Inq* (1995) 79; R. Goldstone, “Justice as a Tool for Peace-Making: Truth Commissions and International Criminal Tribunals”, 28 *N Y U J Int'l L & Pol* (1996) 485; H. Steiner (ed.), *Truth Commissions: A Comparative Assessment* (1997, Harvard). See also D. Dyzenhaus, *Judging the Judges, Judging Ourselves: Truth, Reconciliation and the Apartheid Legal Order* (1998) which focuses on the South African Truth Commission’s hearings into the role of the legal profession during apartheid.

⁴ The Chilean Commission on Truth and Reconciliation was established by Ministry of the Interior Decree No. 355 of 25 April 1990, published in the *Diario Oficial* (9 May 1990). It became known as the ‘Rettig Commission’ after its chairman, the jurist Raul Rettig. On the role of the Chilean Truth Commission, see J. Zalaquett, “Balancing Ethical Imperatives and Political Constraints: The Dilemma of New Democracies Confronting Past Human Rights Violations”, 43 *Hastings L J* (1992) 1425; J. Correa S., “Dealing with Past Human Rights Violations: The Chilean Case After Dictatorship”, 67 *Notre Dame L R* (1992) 1455; R. Quinn, “Will the Rule of Law End? Challenging Grants of Amnesty for the Human Rights Violations of a Prior Regime: Chile’s New Model”, 62 *Fordham L.R.* (1994) 905; M. Ensalaco, “Truth Commissions for Chile and El Salvador: A Report and Assessment”, 16 *Hum Rts Q* (1994) 656.

investigated. And General Pinochet was prepared to co-operate to some extent but only to some extent. Firstly, he insisted, and Aylwin had to accept, that the Commission would be limited to disappearances and would not investigate other forms of human right abuses. So if people had not “disappeared”, this horrible expression which has come to us from Latin America, their human rights violations would not be investigated. The second condition laid down by Pinochet was that none of the hearings would be in public. They would all be behind closed doors. And his third condition was that in no way would the names of the perpetrators ever be made public. So, if one looks at the report of the Truth Commission of Chile⁵, you will find that no perpetrators were named although the victims were. Notwithstanding those deficiencies, the Truth Commission in Chile did help to heal a nation that had been so badly traumatised by the Pinochet military regime, by the terrible human rights violations and abuses for which it was responsible.

When South Africa began to emerge from apartheid, during the transition period people in our society initiated a huge public debate. It was important that the whole of South African civil society contributed to the debate as to what South Africa should do about past human rights abuses. We held two important seminars in Cape Town. We were fortunate in having a very well qualified ambassador from Chile. Ambassador Heine was a civil rights lawyer in Chile for many years. He was not a career diplomat, but he was sent to South Africa fortuitously. And he brought with him to one of the seminars in Cape Town, President Aylwin and members of the Truth Commission. They made a very important contribution to South Africa’s decision to establish a Truth Commission.

At those seminars we heard from some of the victims. Some of us were moved to tears by the stories of the victims. And one of them I recall, and will never forget, was the widow of a black lawyer in South Africa, who was murdered by the security police because he defended people accused of contravening the apartheid laws. He was not involved in politics himself and his widow came to give evidence about the terrible loss to herself, and particularly to her young children, aged eight and ten when their father was murdered. They heard about it, not by any policeman knocking at the door, but on the radio. During the conference I spoke to her and complimented her on her courage in coming to Cape Town from her home, many hundreds of miles away, to tell us her story. And her response to me made a deep impression. She said, “You know, last night is the first night that I’ve been able to sleep through since I heard about my husband’s death.” I asked how she accounted for this and she said, “I don’t know, but I can only put it down to the fact that so many important people from South Africa and from abroad, were interested in hearing my story.” And it was a good illustration, I think, to me, and to people to whom I repeated the story, of the importance of acknowledgement to victims. It was the beginning of that woman’s healing process, the fact that her story was being heard, not only by her

⁵ The Chilean Commission presented its final report, *Report of the Chilean National Commission on Truth and Reconciliation*, in February 1991; see further D. Weissbrodt and P. Fraser, “The Report of the Chilean National Commission on Truth and Reconciliation”, 14 *Hum Rts Q* (1992) 601.

family who knew the story and that she could speak from, what was to her, an important public platform. Many victims know who the perpetrator is and they obviously know what happened to them. But they are not satisfied with that knowledge, they want the acknowledgement that comes with some form of official and public inquiry.

In South Africa, prosecutions for crimes of apartheid were never going to happen. Systematic prosecutions were not feasible for the reason I mentioned – there were simply too many perpetrators and too many victims. More compelling were the political considerations. It was astounding when President de Klerk, the leader of the Government, in control of the armed forces, initiated the transition process. Had he wished, he could have continued the apartheid system for one year, two years, ten years, for all of my life. I have heard that apartheid would last only ten years more; I heard it in the 1950s, in the 1960s, in the 1970s and the 1980s, and it went on. One apartheid leader after another continued with some modifications – one step forward and two steps back - but continued the system. President de Klerk saw apartheid's failure and for that reason decided to begin reforms. It was truly unexpected: it certainly took South Africans and indeed the whole international community by surprise. But it would have been impossible for President de Klerk alone to have agreed to a negotiated transfer of power to a black majority. He needed support from the ruling elite and that support would never have been attained if he had said to his colleagues in his cabinet and in the police and the army, "Not only are we going to hand over power, but we are also going to arrange for Nuremberg style trials and many of us are going to go to prison for the rest of our lives". The transition would never have happened. There would have been no agreement if the leaders of apartheid were inevitably to be put on trial.

President Mandela and the ANC were certainly not prepared to accept a blanket amnesty and opt for the path of amnesia. Had that been the condition laid down by the de Klerk Government, again there would have been no peaceful transition. It was not acceptable to the majority of South Africans that the book on the past should be shut in 1994.

The Truth Commission in South Africa really was a political compromise more than a moral imperative.⁶ The political compromise was that there

⁶ The South African Truth and Reconciliation Commission was based on the final clause of the Interim Constitution of 1993 and was established under s 2(1) of the Promotion of National Unity and Reconciliation Act, No 34 of 1995. The objectives of the Commission are listed in s 3 and include establishing as complete a picture as possible of the causes, nature and extent of the gross violations of human rights which were committed during the period from 1 March 1960 to the cut-off date (10 May 1994); facilitating the granting of amnesty; establishing the fate of victims and granting victims the opportunity to relate their own accounts of the violations perpetrated against them; recommending reparation measures in respect of violations; and compiling a comprehensive report containing recommendations to prevent future violations of human rights. Three committees were also established under the 1995 Act for the purpose of achieving the objectives of the Commission: a Committee on Human Rights Violations, a Committee on Reparation and Rehabilitation, and a Committee on Amnesty. Chaired by Archbishop Desmond Tutu, the Commission began its work in 1995 and published its final report in October 1998 (see n 11 below). On the role of the

would be amnesties, but only in return for full confession, and that incentive-scheme is unique to the South African form of Truth Commission.⁷ There were objections from some of the victims. The families of Steve Biko and Griffiths Mxenge came before our Constitutional Court and argued that the Truth Commission was in violation of certain Constitutional provisions.⁸ Our court, with difficulty, with reluctance and with hesitation, upheld the constitutionality of the Truth Commission, mainly because it had been provided for in the postamble of what was then the interim constitution.

In South Africa, prosecutions were not ruled out as they were in Chile or Argentina. If people did not apply in South Africa for amnesties they should and can be prosecuted if there is sufficient evidence presented to the Attorney General justifying prosecution. High profile prosecutions have been conducted in parallel and simultaneously with the work of the Truth Commission. The Truth Commission was given primacy which meant that if an individual was brought before a court in a criminal prosecution and then applied for amnesty, his trial was almost automatically postponed and suspended, pending the decision of the Truth Commission. Where amnesty was granted, no trial would follow.

When the Truth Commission was set up by Parliament, it was obviously important that it be set up by the first democratic parliament in South Africa's history: a parliament that truly represented the victims of apartheid. It was the representatives of the victims who agreed to the establishment of a Truth Commission. This was of fundamental significance to its moral foundation. It was not the sort of self-amnesty that people like Pinochet granted to themselves, but a decision taken almost unanimously. Although there was opposition to it from the extreme white right-wing party in Parliament, the Truth Commission had the overwhelming support of the members of the democratic South African Parliament. That fact is often forgotten in the debate on the moral justification of the South African Truth Commission.

When it was first established, its supporters were extremely nervous. Would it work? Would perpetrators come forward? Would victims come forward and give evidence of what had happened to them? Would any significant figures apply for amnesty and make full confessions? This was by no means inevitable. I headed the Commission of Inquiry into Violence in South Africa between 1991 and 1994 and, fortunately, we were able at least to scratch the surface of the police and military involvement in serious

South African Truth Commission, see I. Liebenberg, "The Truth and Reconciliation Commission in South Africa", 11 *S Afr P L* (1996) 123; and J. Sarkin, "The Trials and Tribulations of the South African Truth and Reconciliation Commission", *S Afr J Hum Rts* (1996) 617. On the operation of the three committees, see the Commission's website at: www.truth.org.za/.

⁷ On the debate over amnesties and indemnities, see P. Parker, "The Politics of Indemnities, Truth Telling and Reconciliation in South Africa: Ending Apartheid Without Forgetting", 17 *Hum Rts L J* (1996) 1; and E. McCarthy, "South Africa's Amnesty Process: A Viable Route Toward Truth and Reconciliation?", 3 *Mich J Race & L* (1997) 183.

⁸ See *Azanian Peoples Organisation (AZAPO) and Others v President of RSA and Others*, 8 BCLR (1996) 1015 (Constitutional Court), Deputy President Mahomed speaking for the majority of the Court.

criminality, even during the negotiating period. I have no doubt that, without the knowledge of the then President de Klerk (that he should have known, it may well be, but that he did not know I have little doubt) elements in the military, and in the police, who did not want the negotiations to succeed, who did not want a black government and who would have preferred apartheid to continue, were sabotaging the whole negotiation process by committing the most terrible criminal offences. Exposing the involvement of senior members of the military and the leaders of the police in criminal activity certainly helped to pave the way for the Truth Commission, because the denials had already begun. We heard for years in South Africa, particularly in the 1990s, that these allegations of murders and torture were untrue; that they were not committed by the police; that this was propaganda put out by the African National Congress and other liberation movements. We were told we did not need a Truth Commission because there was no truth unknown. The disclosure of some of the truth, in 1993 and 1994, was sufficient to pull the rug from under the feet of the people who were putting forth these denials, and they were forced, in the light of those revelations, to agree to and to participate in establishing the Truth Commission.

The most recent Truth Commission report, that of Guatemala is interesting too. It is a very different form of Truth Commission, a Truth Commission that was set up by agreement when the military regime came to an end.⁹ The United Nations was asked to appoint the chairman of their Truth Commission, and they appointed an eminent German international lawyer, Professor Tomuschat from Humboldt University. He sat with and investigated alongside two local Guatemalan lawyers. The agreement was that this Commission, like the Chilean Commission, would not name names. They would not hold hearings in public. But nonetheless their report, which was published two or three months ago, has been very important already for change in Guatemala. It exposed the involvement and complicity of former leaders of Guatemala, who denied their involvement, and it also disclosed, to the credit of the United States, the covert assistance and finance which the governments of the military in Guatemala received from the CIA, who were aware of the criminal activities of those governments.¹⁰

⁹ The decision to establish a truth commission in Guatemala dates back to June 1994 when, as part of the negotiations to end the conflict and under intense international pressure, the Guatemalan Government and the guerrilla movement known as the 'Guatemalan National Revolutionary Union' (URNG) agreed to the formation of a truth commission; see *Accord on the Establishment of the Commission to Clarify Past Human Rights Violations and Acts of Violence that have caused the Guatemalan Population to Suffer*, signed in Oslo, 23 June 1994 (UN doc. A/48/954-S/1994/751, 1 July 1994). Although agreement was reached in 1994 to establish a truth commission, it was not to begin work until the parties had signed a final peace accord. This occurred in December 1996 and was followed by the establishment of the 'Historical Clarification Commission' which formally began its work in August 1997.

¹⁰ The final report of the Guatemalan Historical Clarification Commission was presented to President Alvaro Arzu and declared public by the Secretary-General of the United Nations, Kofi Annan, in accordance with the procedures contained in the 1994 Accord (see n 9 above) on 25 February 1999. In his inaugural address on 14 January 2000, the new President of Guatemala, Alfonso Portillo, reaffirmed his

Public interest is served by the search for truth and its public inscription. Although there will always be those who attempt denials and revisionist accounts of the past, their impact can be lessened. In South Africa, without the Truth Commission, there would have been at least two histories as there are three in the former Yugoslavia. Denials issued by the apartheid government would have been believed, certainly by most white South Africans. White South Africans would have wanted to believe the denials because they would have alleviated their guilt in respect of what they knew, or what they did not know, what they could have done and what they did not do. That belief cannot stand up against the evidence presented by the Truth Commission. Disclosure is relevant too in preventing a recurrence of atrocities because much of the evidence presented by the South African Truth Commission has ensured that certain people who would have remained in official positions, in the police and the army in particular, have been removed from office.

In the former Yugoslavia, the work of the War Crimes Tribunal, too, has lessened the impact of denials and in some ways may have helped prevent recurrences, at least in Bosnia. The Bosnian Serb army denied the massacres of 8,000 men and boys outside Srebrenica in July 1995. It was only when evidence was obtained from one of the Bosnian Serb soldiers who made a confession to us, that one of the mass graves was found and exhumed. It contained the bodies of boys and adult men, each had been killed by a single bullet wound to the head. Before the exhumation, the Bosnia Serb army had denied the massacre, had said that if there were graves, they were of those who had died in the war. Here too public disclosure helped to prevent denial.

In East Germany, similar problems were faced at the end of the communist era, when it was decided that the Stasi files should be made public and administered by the Gauck authority. To this day, if a former Eastern German wants any official position in government, at state or federal level, the Gauck administration must certify that the files do not incriminate the individual concerned. There too the exposure of the truth has helped to ensure that people who should not be in positions of authority are removed.

Disclosure averts collective guilt. Collective guilt has been the fuel for the tragedies in Rwanda and the former Yugoslavia. When blame is levelled at an entire people indiscriminately, individuals such as Karadzic and Mladic can take advantage and bring about the calamities to their country and people.

The advantage of a Truth Commission too, as opposed to prosecutions, is its broad focus. In South Africa, the Truth Commission has a mandate to expose human rights abuses during a period of 34 years, between 1960 and 1994. It would take many years before the same amount of information could be exposed by prosecutions. The Truth and Reconciliation Commission has been able to condense its investigations into a shorter period because of the response it received. Little could anyone have anticipated that there would be over 8,000 applications for amnesty and over 21,000 victims coming to tell their story, covering this period. And the huge five-volume report of the

commitment to establish a mechanism to follow up on the recommendations of the Commission.

Truth Commission is a testimony to that.¹¹ The history of apartheid which is recorded in the report of the Truth Commission, I think, will become a shared history of black and white South Africans. The process has avoided collective guilt because the Truth Commissioners themselves came from all parts of our country, black and white people served on the Commission. The victims too came from the white community and the black community and the Asian community. So it was not a question of whites or blacks judging blacks or whites. It was South Africans coming together to judge fellow South Africans. It was this shared history that will avoid the collective guilt.

Amnesties are not essential to Truth Commissions. South Africa made amnesties serve as an incentive-scheme. Without the amnesties a lot of the truth would not have emerged. Yet the impressive fact that there were over 8,000 amnesty applications obscures one troubling reality. About 55% of the applications came from people who had already been convicted to – people in prison who really had little to lose by confessing. Fortunately, the Truth Commission had an efficient investigation department and many amnesties have been refused because one of the conditions for the amnesty, a full disclosure, was not met. Many families opposed particular amnesties and the Truth Commission itself brought forth evidence in contradiction of some of the disclosures that were made.¹² There is talk of a Truth Commission now in Bosnia. Many individuals and NGOs, not government, are talking about the need for a Truth Commission in addition to the United Nations Tribunal. And if the people of Bosnia want it, I have no doubt they should be encouraged. I do not agree with some of the views expressed by people working for the UN War Crimes Tribunal that a Truth Commission would conflict with the work of the Tribunal. I think the South African experience has shown that there is no contradiction, as long as one institution is given primacy, and where you have a Security Council Tribunal established, clearly it would have to be given primacy in the Bosnian situation. But if victims were encouraged to come forward, and to state in public what happened to them in Bosnia (and there were victims on all sides in Bosnia) and to tell their stories, I have little doubt that in the experiences they relate, they would only assist the prosecutions by the International Criminal Tribunal.

This whole concept of “Truth Commissions” which started in South America, moved to Africa, and was incorporated within Germany is

¹¹ The *Final Report of the Truth and Reconciliation Commission* was presented to President Mandela on 29 October 1998, the full text of which can be accessed via the Commission’s website at: www.truth.org.za/. Volume One (twelve chapters) is an introductory volume which describes the working methods of the Commission; Volume Two (seven chapters) addresses the commission of gross violations of human rights on all sides of the conflict; Volume Three (five chapters) focuses on gross violations of human rights from the perspective of the victim; Volume Four (ten chapters) investigates the nature of the society in which gross violations of human rights occurred and reports on a series of ‘institutional hearings’; and Volume Five (nine chapters) contains the conclusions and recommendations of the Commission (including the minority position of Commissioner Wynand Malan).

¹² For transcripts of the decisions on amnesty, see the Commission’s website at: www.truth.org.za/.

composed of so many variables that it is not a tidy process one can simply import or export. It is certainly of significance to any society moving from a period of violence, from a period of ethnic division, or religious division, to a period of democracy and reconciliation. But the extent to which a particular society can employ the process can only be determined by the society itself. The choice really must depend on the history, on the culture, on, above all, the prevailing politics. I hope that the South African experience will be relevant to Ireland, I do not know that it is. One thing I do know is that South Africa has received so much from the international community. We have received so much advice, so much assistance, and more importantly, the political opposition to apartheid, without which apartheid would have not come to an end. And I can assure you only of one thing as a South African, that any assistance that you may need from South Africa will be gladly and open-heartedly given to you.

DISADVANTAGE AND DISCRIMINATION: THE EMERGING JURISPRUDENCE OF THE SOUTH AFRICAN CONSTITUTIONAL COURT

*Evadné Grant, Law Department, Oxford Brookes University and
Joan G. Small, Law Department, City University*

INTRODUCTION

In most countries of the world, and in much of the existing international human rights law, equality is recognised in the Aristotelian sense.¹ Aristotle's view was that "justice considers that persons who are equal should have assigned to them equal things".² Consequently, "there is no inequality when unequals are treated in proportion to the inequality existing between them".³ Laws which seek to draw distinctions between groups of persons and treat them differently must first, identify those who are alike correctly so as to draw appropriate distinctions between persons, and second, provide for appropriately different treatment for those who are not alike. Aristotle's definition provides the test for laws that do not offend the principle of equality, but it is formal, without substantive content: there are no criteria by which to assess appropriate distinctions and appropriately different treatment.

The (virtually) universal approach to equality in law is to declare the equality of all persons, or to imply, from the drafting of the human rights provisions, that this equality is an *a priori* fact, from which the legal guarantee of equal treatment or protection before or under the law flows and from which follows the prohibition of discrimination on specified or unspecified grounds. Equality norms in law therefore, generally, follow the Aristotelian approach, while supplying the categories or classes which are not appropriate bases on which to draw distinctions from which legal differences will follow. They therefore give some normative content to the Aristotelian criteria for those who are alike, but that content is not substantively elaborated. While the model relating to equality presumes equality of fact, those often more specific provisions outlawing discrimination are usually expressly premised on the desire to redress perceived injustices or disadvantage being experienced by a particular group or groups within a society or community. Women or ethnic or religious minorities, for example, are identified as participating to a disproportionately and unfairly small degree in such areas as employment and educational opportunities.⁴ Anti-discrimination laws specifically created for the relief of such disadvantage generally provide punitive measures where active prejudice results in disadvantaged treatment,

¹ See Aristotle's *Nicomachean Ethics*, (trans. Ross, *World's Classic Series* 1980).

² *The Politics of Aristotle* (trans. Barker, 1946) Book III, xii, 1282b.

³ Book V, i, 1301a.

⁴ See for example, the White Paper *Equality for Women* which preceded the Sex Discrimination Act 1975 in the United Kingdom. The justification for the legislation was stated as follows: "The status of women in society, the disabilities and disadvantages imposed on women and their consequences, are social questions. They are legitimate subjects of the public interest and are appropriate matters for government action." Cmnd 5724 (1974), para 3.

as well as measures designed to relieve that more subtle “institutionalised” discrimination. The logic underlying these latter measures is compensation for the victims of the prior system, which acts both remedially and as a deterrent to any repeated injustices in the future. Therefore, however the courts may interpret them, the conceptual foundation for such provisions is that it is the effect on the victims with which the law is concerned, not the conduct of the perpetrators.

It is the legal recognition of the conceptual foundations of such legislation which has challenged the formal, public law model of equality laws. Ironically, though specifically designed to alleviate disadvantage, those anti-discrimination laws are often drafted using the same, formal criteria for claims as equality laws which adopt the simple legal statement of the Aristotelian ideal. Witness the United Kingdom Sex Discrimination Act 1975, for example: the White Paper introducing it was titled “Equality for Women”, and virtually all Parliamentary debate concerned the alleviation of women’s disadvantaged status as the legislative goal. But, despite the rhetoric accompanying its enactment, the legislation adopts a formal concept of equality between the sexes, rather than specific measures designed to cure the identified social evils of women’s status. The case law under the Act has demonstrated the failure of the provisions to cure many of the problems which are most pressing for women, precisely because they are incapable of argument under the provisions which are claimed to establish equality in law – and the resulting equality of fact. The gulf between the justification for the measures and the actual rules created is certainly not unique among legislative provisions aimed at the question of discrimination: but it serves to highlight the difficult concept of equality in law and the idea of the alleviation of discrimination in the name of equality.

Since discrimination and equality have been conceived as “two sides of the same coin”, it is problematic to introduce into the concept of discrimination the idea of the alleviation of disadvantage without also challenging the very notion of equality in law. Central to the re-evaluation of the meaning of equality (or discrimination) is the question of the purpose or purposes of equality law, and it is in relation to that purpose or those purposes that the concept of disadvantage – and its precise legal status – is beginning to be considered. This re-evaluation of disadvantaged status is proving particularly intense and controversial under the new South African Constitution.

The questions of ensuring equality and redressing past disadvantage have particular resonance and relevance in Northern Ireland. Like South Africa, Northern Ireland has been engaged in a process of constitutional change against a background of historical and entrenched inequality. The Good Friday Agreement identified equality as central to the settlement process. The subsequent Northern Ireland Act of 1998, which forms the legislative basis for a new constitutional settlement in Northern Ireland makes provision for a number of measures aimed at addressing the problem of discrimination. Most significant among these is a statutory duty placed on public authorities to carry out their public duties having due regard to the need to promote equality of opportunity between persons and groups in a number of categories, which include religious belief, political opinion and racial groups

amongst others.⁵ In addition, the equality requirements of the European Convention on Human Rights will begin to apply in Northern Ireland when the Human Rights Act 1998 comes into operation in October 2000. The effect of these measures for Northern Ireland has been rightly assessed as one of mainstreaming equality issues, in the sense that public authorities will be required to engage in an analysis of the impact of their decisions in relation to the greater goal of equality – in law and in fact – within the jurisdiction.⁶ It is therefore all the more important that such public authorities have an understanding of equality (and discrimination) which adequately serves the purposes of mainstreaming. While Northern Ireland has to some extent engaged on an analysis of those issues which relate to substantive equality,⁷ it is suggested that the South African experience in re-evaluating equality and discrimination within the context of disadvantage provides further useful insights.

In light of South Africa's history of legally enforced and socially entrenched racial discrimination, it is hardly surprising that equality is one of the fundamental principles which has shaped the Constitution and will continue to do so.⁸ In the words of one of the judges of the Constitutional Court, "the Constitution is an emphatic renunciation of our past in which inequality was systematically entrenched."⁹ Like most constitutional equality guarantees around the world, the equality provisions of the Bill of Rights follow the Aristotelian model. But, given the history of apartheid and the grave disadvantage suffered by the majority population, the very purpose of the equality clause has been controversial. Against this background, the Constitutional Court has commenced the Herculean task of contextualising the notion of equality and developing a framework for the adjudication of discrimination claims.¹⁰

⁵ See McCrudden, "Mainstreaming Equality in the Governance of Northern Ireland" (1999) 22 *Fordham International Law Journal* 1696 for a detailed discussion of the background and new legislative provisions.

⁶ McCrudden, note 5 at 1698.

⁷ Particularly in relation to the Fair Employment (Northern Ireland) Act 1989. See Hepple, "Discrimination and equality of opportunity – Northern Irish lessons" (1990) 10 *OJLS* 408 and McCrudden, "Affirmative Action and fair participation: interpreting the Fair Employment Act 1989" (1992) 21 *Industrial LJ* 170.

⁸ In *President of the Republic of South Africa and Another v Hugo* 1997 (4) SA 1 (CC), Justice Kriegler described equality as the central "organising principle" of the Constitution (at para 74).

⁹ Justice O'Regan in *Brink v Kitshoff NO* 1996 (4) SA 197 (CC), 1996 (6) BCLR 752 (CC) at para 33.

¹⁰ To date there have been eight cases in which discrimination was the only or one of the main issues: *Brink v Kitshoff NO* 1996 (4) SA 197 (CC), 1996 (6) BCLR 752 (CC); *Prinsloo v Van der Linde and Another* 1997 (3) SA 1012 (CC), 1997 (6) BCLR 759 (CC); *President of the Republic of South Africa and Another v Hugo* 1997 (4) SA 1 (CC), 1997 (6) BCLR 708 (CC); *Harksen v Lane NO and Others* 1998 (1) SA 300 (CC), 1997 (11) BCLR 1489 (CC); *Larbi-Odam and Others v Member of the Executive Committee for Education (North West Province) and Another* 1998 (1) SA 745 (CC), 1997 (12) BCLR 1655 (CC); *Pretoria City Council v Walker* 1998 (2) SA 363 (CC), 1998 (3) BCLR 257 (CC); *The National Coalition for Gay and Lesbian Equality and Another v Minister of Justice and Others (National Coalition No 1)* 1999 (1) SA 6 (CC), 1998 (12) BCLR 1517 (CC); *The National Coalition for Gay and Lesbian Equality and Others v Minister*

In one of the earliest equality cases before the Constitutional Court, *Brink v Kitshoff*,¹¹ O'Regan J, delivering the unanimous judgment of the Court stated:

“Section 8¹² was adopted . . . in the recognition that discrimination against people who are members of disfavoured groups can lead to patterns of group disadvantage and harm. Such discrimination is unfair: it builds and entrenches inequality amongst different groups in our society.”¹³

This dictum is often cited by commentators who argue that the reversal of systemic discrimination and patterns of group disadvantage are the central characteristics of substantive equality. Equality is understood as remedial in nature, associated with the protection of those who have suffered from historical and systemic disadvantage and the creation of the circumstances in which equality of outcome is made possible for such applicants.¹⁴ Furthermore, it is contended, if the primary aim of the equality clause is remedial and transformative, its protection is to be restricted to those who have suffered disadvantage in the past. This, it is argued, requires the Constitutional Court to interpret equality substantively in the light of the real social and economic circumstances of the victims of discrimination¹⁵ and to allot to the notion of historical disadvantage a key role in determining discrimination.¹⁶ But, while the Constitutional Court has, from the outset, agreed on the need to develop a substantive equality jurisprudence, the role of disadvantage in the adjudication of discrimination claims has been much more problematic.

Non-discrimination in the South African Bill of Rights¹⁷

Non-discrimination is provided for in section 9(3)¹⁸ of the Constitution, which states:

of Home Affairs and Others (National Coalition No 2) Judgment of 2 December 1999, as yet unreported.

¹¹ Above note 10.

¹² The case concerned the interpretation of the equality clause in the interim Constitution (Act 200 of 1993). Although there are textual differences between the equality clauses in the interim and final Constitutions, the Constitutional Court held in *National Coalition for Gay and Lesbian Equality (No 1)* (above note 10) that the equality jurisprudence developed in relation to the interim Constitution remains applicable (at para 15).

¹³ At para 42.

¹⁴ See Albertyn and Goldblatt, “Facing the challenge of Transformation: Difficulties in the development of an indigenous jurisprudence of Equality” (1998) 14 *South African Journal on Human Rights* 248 at 256-7.

¹⁵ See for example, Albertyn and Kentridge, “Introducing the Right to Equality in the Interim Constitution” (1994) 10 *South African Journal on Human Rights* 149; Davis, “Equality and Equal Protection” in Van Wyk, Dugard, De Villiers and Davis (eds) *Rights and Constitutionalism: The New South African Legal Order* 196.

¹⁶ See Albertyn and Goldblatt, note 14.

¹⁷ The Bill of Rights is Chapter 2 of the Constitution, Act 108 of 1996.

¹⁸ Section 9(1) provides for equal protection of the law. It has been the subject of argument in favour of a substantive law which focuses on disadvantage, but the Constitutional Court has rejected this approach, firmly placing the notion of

“The state may not unfairly discriminate directly or indirectly against anyone on one or more grounds, including race, gender, sex, pregnancy, marital status, ethnic or social origin, colour, sexual orientation, age, disability, religion, conscience, belief, culture, language and birth.”

That such an extensive list of prohibited grounds has been specified is not surprising in light of the history and the nature of the constitution-making process. The use of the term “including” means that the list is not exhaustive, permitting applications on grounds other than those stated. Section 9(4) further extends the potential reach of the provision in providing for horizontal application thereof.¹⁹ Because of the peculiar formulation prohibiting unfair discrimination rather than discrimination *per se*, the Constitutional Court has established a two stage analysis. At the first stage the question is whether the acts or legislative provisions challenged by the complainant amounted to discrimination. The second stage is concerned with the question of unfairness.²⁰ It is significant for the second stage that, in relation to the specified grounds of discrimination, section 9(5) reverses the burden of proof of unfairness, placing it on the alleged perpetrator of discrimination, rather than the victim.²¹ The division of the enquiry into two stages thus has both procedural and conceptual implications: the analysis at the first stage may have a decisive impact on the way in which the Court approaches the enquiry at the second stage which may be crucial to the success or failure of applications under section 9(3).²²

Stage 1: The meaning of discrimination:

Clearly, differentiation or different treatment of individuals or groups of individuals does not, of itself, constitute discrimination. While discrimination at root means differentiation, it carries opprobrious connotations which distinguishes it from acceptable and essential differentiation. But the use of the term ‘unfair’ to qualify discrimination in section 9(3) raises the question whether, in this context, discrimination has been stripped of its pejorative signification.²³ The Constitutional Court has, however, firmly

disadvantage and its legal status within an analysis under section 9(3). See Small and Grant, “Equality and Non-discrimination in the South African Constitution” (2000) 4 *IJDL* 47. Section 9(2) deals with affirmative action, but has not yet been considered in the Constitutional Court. Accordingly, arguments here are restricted to section 9(3).

¹⁹ The subsection also provides that national legislation is to be enacted to give effect to horizontal application. The Promotion of Equality and Prevention of Unfair Discrimination Act 2000, the aim of which is to give effect to s 9(4), was passed in January 2000. The legislation is controversial, opponents arguing that it does not strike a proper balance between the right to equality and other important rights in the Bill of Rights. See *Financial Times* 27 October 1999 and 27 January 2000.

²⁰ See *Harksen* (note 10) at para 45.

²¹ “Discrimination on one or more of the grounds listed in subsection (3) is unfair unless it is established that the discrimination is fair.”

²² If at the end of the two-stage enquiry it is held that a provision or act is unfairly discriminatory, it may still be justified in terms of the limitation clause, section 36 of the Constitution, as “reasonable and justifiable in an open and democratic society based on human dignity equality and freedom”.

²³ See Davis, (note 15) at 208.

rejected any suggestion that discrimination had become a neutral term because of the addition of 'unfair', stating in *Prinsloo v Van der Linde*:

"The proscribed activity is not stated to be "unfair differentiation" but is stated to be "unfair discrimination". Given the history of this country we are of the view that 'discrimination' has acquired a particular pejorative meaning relating to the unequal treatment of people based on attributes and characteristics attaching to them. . ." (emphasis in the original).²⁴

According to the Court, two types of discrimination are provided for in section 9(3), namely discrimination on a specified ground and discrimination on an unspecified ground. In its analysis, differentiation on a specified ground is assumed to be discriminatory, while differentiation on a ground not specified must be examined by the court to determine, on a case by case basis, whether the provision or action is discriminatory.²⁵

What is it about differentiation on the specified grounds which renders such action immediately discriminatory? According to Goldstone J in the leading case of *Harksen v Lane*:

"What the specified grounds have in common is that they have been used (or misused) in the past (both in South Africa and elsewhere) to categorise, marginalise and often oppress persons who have had, or who have been associated with those attributes or characteristics. These grounds have the potential when manipulated, to demean persons in their inherent humanity and dignity. . . Section 8(2) seeks to prevent the unequal treatment of people based on such criteria which may, amongst other things, result in the constitution of patterns of disadvantage such as has occurred only too visibly in our history."²⁶

In order to prevent the perpetuation of inequality, use of any of the specified grounds as the basis for differentiation is thus in itself suspect. Although the same historical experience which informed the selection of the fourteen grounds will also have shown that it is usually particular groups, such as women or members of particular racial groups who have been disadvantaged by differentiation, section 9(2) itself does not distinguish such groups as being more deserving of protection than others. As with other provisions in the Aristotelian mould, it is couched in neutral terms. Thus, for example, it is discrimination on the ground of race which is prohibited, rather than discrimination against particular racial groups who have been disadvantaged in the past. In the case of *Pretoria City Council v Walker*²⁷, the applicants alleged that the City Council had discriminated against them on racial grounds. But the applicants themselves, white residents of a relatively affluent suburb of Pretoria, had clearly not been disadvantaged in the past, especially not in relation to the provision of services and amenities which

²⁴ *Prinsloo* (note 10) at para 31 .

²⁵ *Prinsloo* at para 28; see also *Harksen* (note 10) at para 46.

²⁶ *Harksen* at para 49.

²⁷ See note 10.

was the particular issue at stake in the case. Nonetheless, the Constitutional Court accepted that differentiation between black and white residents constituted discrimination without further ado²⁸. In *President of the Republic of South Africa and Another v Hugo*²⁹, the applicant, a male prisoner, alleged sex discrimination. Although it is women who have suffered historically in relation to men, the Constitutional Court accepted that the applicant had suffered discrimination merely because of different treatment of men and women prisoners with regard to early release.³⁰ In this conception, discrimination becomes a legal status born from a factual situation of differentiation based on one of the stated categories.

But past disadvantage is not the only factor which is highlighted by the Court in *Harksen* as common to the specified grounds. The use of such grounds for purposes of categorisation or differentiation also has the “potential to demean persons in their inherent humanity and dignity”.³¹ It is this factor which the Constitutional Court has identified as the central to the determination of discrimination on an unspecified ground, formulating the test as follows:

“There will be discrimination on an unspecified ground if it is based on attributes or characteristics which have the potential to impair the fundamental dignity of persons as human beings, or to affect them adversely in a comparably serious manner.”³²

Just what may affect someone in a manner comparably serious to the impairment of fundamental dignity is yet to be fully worked out.³³ Although historical disadvantage is not explicitly identified as a factor in determining discrimination on an unspecified ground, it may nonetheless play a role in practice. In the case of *Larbi-Odam and Others v Member of the Executive Council for Education (North West Province) and Another*³⁴ a discrimination application was brought by teachers denied permanent appointments and other opportunities because they were non-citizens. In concluding that differentiation on the ground of citizenship had the potential to impair the fundamental human dignity of the applicants, Mokgoro J, in whose judgment the rest of the court concurred, said:

“[The] general lack of control over one’s citizenship has particular resonance in the South African context, where individuals were deprived of rights or benefits ostensibly on the basis of citizenship, but in reality in circumstances where citizenship was governed by race.”³⁵

²⁸ At para 32. The approach of the Constitutional Court has been criticised by Jagwanth, “What is the Difference? Group Categorisation in *Pretoria City Council v Walker* 1998 (2) SA 363 (CC)” (1999) 15 *South African Journal on Human Rights* 200.

²⁹ See note 10.

³⁰ At para 33. The application failed however, because it was held that the discrimination was not unfair (para 47).

³¹ At para 49.

³² *Harksen* at para 46. See also *Prinsloo* (note 10) at para 33.

³³ See *Walker* (note 10).

³⁴ See note 10.

³⁵ At para 19.

Thus the court recognised that black non-citizens, in particular, had been historically disadvantaged, and this is used as a factor in determining the potential to impair the dignity of the applicant, thereby to establish discrimination.

Determining whether an applicant's claim is based on a specified or unspecified ground has important procedural implications because of the reversal of the burden of proof provided for in section 9(5). The case of *Walker*³⁶ is illustrative. The applicants were residents of a suburb of Pretoria, which had, under apartheid, been designated as "white". Their discrimination claim was based on the fact that residents in previously "white" areas were paying more for electricity and water services than residents in previously "black" areas, and that a different policy in relation to non-payment operated in relation to the two areas. The majority of the Constitutional Court concluded that this was a case of indirect discrimination, based on race.³⁷ This being a specified ground, the applicants had two very important advantages which they would not have had if the majority of the court had agreed with the minority judgment of Sachs J, that the ground of discrimination at issue was geographical location.³⁸ First, it was not necessary for the applicants to prove discrimination because they were able to rely on a specified ground which permits the court to assume discrimination. Secondly, because of the reversal of the usual burden of proof provided for in subsection 5, the applicants were in the fortunate position of not having to prove unfairness either. Instead it was for the defendants to prove that the differentiation was not unfair.

The decision as to the applicable grounds of discrimination is also significant for the substance of the enquiry into unfairness at the second stage. The question of unfairness arises only in relation to the specific ground established as the basis for discrimination at the first stage. Thus for example in *Walker*, because the majority judgment held that the defendant had discriminated against the applicant on the grounds of race, it was the unfairness of race discrimination which was at issue in the second stage. By contrast the focus of the unfairness analysis in the minority judgment of Sachs J related to discrimination on the grounds of geographical location. As the judgment shows, the evidence, scope of enquiry and outcome may be very different in the case of discrimination on the grounds of geography than it is in relation to race discrimination. Moreover, the extent to which disadvantage features as a relevant consideration may differ depending on the ground of discrimination being analysed. In *Harksen*³⁹, O'Regan J entered a strong dissent⁴⁰ to the majority judgment of Goldstone J. The issue in the case was whether certain provisions of the Insolvency Act⁴¹ which impacted on spouses of insolvents offended the equality clause. Whereas the majority decision categorised the applicable ground of discrimination in the case as between spouses of insolvents as compared with other persons who

³⁶ See note 10.

³⁷ This was the first case in which indirect discrimination was considered by the Constitutional Court, see para 30-33.

³⁸ At para 118.

³⁹ See note 10.

⁴⁰ Supported by Madala J and Mokgoro J.

⁴¹ Act 24 of 1936 as amended.

had dealings with insolvents, the minority identified marital status as the primary basis of differentiation. This difference in approach had no effect on the burden of proof in relation to unfairness because marital status was not at that time a specified ground.⁴² But, as in *Walker*, the difference in approach did have a profound impact on the scope and substance of the unfairness enquiry, in which the relevance of disadvantage played a key part.

Such difference in opinion regarding the basis of the claim at the first stage is particularly problematic in cases in which applicants wish to rely on more than one ground. Section 9(3) refers to discrimination “on one or more grounds”, clearly contemplating applications based on either single or multiple grounds. But, since discrimination is assumed if based on a specified ground, it is only necessary to show differentiation on one of the specified grounds in order to activate the assumption of discrimination and to shift the onus of proof regarding fairness onto the defendant. In light of this, the Court has ruled that if discrimination on a single specified ground has been established, it is unnecessary to investigate additional grounds. In *Brink*,⁴³ for example, the applicant relied on two grounds of discrimination, sex, a specified ground, and marital status, which was not a specified ground at the time⁴⁴. In giving the unanimous judgment of the court, O’Regan J, said that since sex is a specified ground it was unnecessary to consider whether marital status would be a ground, “. . . [i]t is sufficient that the disadvantageous treatment is substantially based on one of the listed prohibited grounds. . .”.⁴⁵ The case thus proceeded to the second stage on the basis that the only issue to be determined was the unfairness of sex discrimination. Although the Court actually referred to the fact that discrimination on a particular ground may be compounded by discrimination on overlapping grounds⁴⁶, this did not actually play a role at the second stage because there was no finding of discrimination based on marital status. If the Court were to persist in this approach, there would be little point in relying on multiple grounds especially if any additional grounds are unspecified. The knock-on effect at the second stage is that circumstances which may be relevant to the overall experience of the applicant, such as disadvantage stemming from different types of discrimination may not be taken into account. The undesirable effect of this may well be an impoverished and one-dimensional equality jurisprudence which fails to come to grips with the real experience of the victims of discrimination.

There are, however, signs of a better understanding of the impact of discrimination on overlapping grounds in the most recent discrimination decision of the Constitutional Court, *The National Coalition for Gay and Lesbian Equality and Another v Minister of Home Affairs and Others*

⁴² Marital status is now a specified ground. If the case had been heard after the 1996 Constitution had come into effect, the impact of the difference in approach would have been amplified since the burden of proof would also have differed.

⁴³ See note 10.

⁴⁴ The case was decided under the 1993 Interim Constitution. Marital status is now a specified ground under s 9(3) of the Constitution.

⁴⁵ At para 43. In *Hugo* (note 10) the Constitutional Court similarly ignored one of the grounds, since a single ground, sex, was held to be sufficient in order to prove breach of the non-discrimination provision (at para 33).

⁴⁶ At para 44.

(*National Coalition No 2*)⁴⁷. The case concerned the constitutionality of immigration laws facilitating the immigration of spouses of South African citizens, which the respondents refused to apply to same-sex partners. Although it was possible to view the application as one based on the ground of sexual orientation alone, Ackermann J, in an unanimous judgment said that the better view was that the discrimination in question constituted “overlapping or intersecting discrimination on the grounds of sexual orientation and marital status”.⁴⁸ In coming to this conclusion the judgment explicitly recognised that an approach which is substantive and contextual requires account to be taken of the interrelationship between grounds of discrimination and the full impact of discrimination on multiple grounds. In this particular case the burden of proof was not affected by the use of joint grounds because both sexual orientation and marital status are specified grounds, but the determination of unfairness proceeded on the basis of an investigation of the effect of discrimination on the applicants, both in relation to sexual orientation and in relation to their relationships. This suggests a growing awareness on the part of the court of the problem of overlapping or intersecting discrimination and willingness to engage with the issues in a more careful and thoughtful manner.

It may be argued that one of the strengths of the two-stage enquiry implicit in the South African provision is that formal equality is determined at the first stage, while the unfairness enquiry at the second stage provides the courts with a separate opportunity to assess substantive equality.⁴⁹ Thus, it may be contended that the absence of any real opportunity to consider disadvantage at the first stage is not significant, since it has greater relevance to the enquiry into the substantive impact of the impugned measure which takes place at the second stage. However, although this provides a neat explanation of the peculiar formulation and application of the provision, the uncritical acceptance that the two stages perform separate functions fails to take account of the potential impact of decisions at the first stage on the second. Since the first stage of the enquiry determines the scope and substance of the enquiry into unfairness at the second stage, decisions made at the first stage have the potential to limit the substantive analysis. As the discussion of the cases shows, this is an issue which has not been discussed by the Constitutional Court: indeed there is little indication in the cases that the Court fully appreciates its significance.

Stage 2: The meaning of Unfairness

In *Brink*, O’Regan J asserted that the equality clause was adopted in the recognition that the disadvantage to disfavoured groups by discriminatory treatment is unfair because it builds and entrenches inequality amongst different groups.⁵⁰ This dictum is often invoked to support the argument that the requirement of unfairness was specifically inserted in the equality clause in order to provide a means of distinguishing mere discrimination, which may affect both advantaged and disadvantaged applicants, and discrimination

⁴⁷ See note 10.

⁴⁸ At para 40.

⁴⁹ See Albertyn and Kentridge, (note 15) at 161-2.

⁵⁰ At para 42.

against historically disadvantaged groups, which is unfair. Accordingly, so goes the argument, discrimination is unfair when perpetrated against persons or groups who have suffered historical and systemic disadvantage.⁵¹ But the subsequent cases of *Hugo*⁵² and *Prinsloo*,⁵³ decided on the same day, provide a more detailed analysis of the unfairness requirement than *Brink*, an analysis, moreover, in which the notion of historical disadvantage is not given an exclusive or even controlling role. In *Hugo*, the court stated its understanding of the purpose of the equality clause as follows:

“The prohibition on unfair discrimination in the interim Constitution seeks not only to avoid discrimination against people who are members of disadvantaged groups. It seeks more than that. At the heart of the prohibition of unfair discrimination lies a recognition that the purpose of our new constitutional and democratic order is the establishment of a society in which all human beings will be accorded equal dignity and respect regardless of their membership of particular groups. The achievement of such a society in the context of our deeply inegalitarian past will not be easy, but that that is the goal of the Constitution should not be forgotten or overlooked.”⁵⁴

In the evolving approach of the Constitutional Court, it is the impact on the victims of discrimination which is to be examined in order to establish unfairness. This impact is to be evaluated on the basis of three key considerations which were confirmed and elaborated in *Harksen v Lane NO*⁵⁵, now considered to be the leading authority.⁵⁶ According to *Harksen* the three factors are:

- “(a) the position of the complainants in society and whether they have suffered in the past from patterns of disadvantage, whether the discrimination in the case under consideration is on a specified ground or not;
- (b) the nature of the provision or power and the purpose sought to be achieved by it. . . .
- (c) with due regard to (a) and (b) above and any other relevant factors, the extent to which the discrimination has affected the rights or interests of complainants and whether it has led to an impairment of their fundamental human dignity or constitutes an impairment of a comparably serious nature.”⁵⁷

Clearly, in the view of the Constitutional Court, historical disadvantage has a role to play in determining unfairness. But it is not the only consideration, or even the most important consideration. It is one of three identified factors in

⁵¹ *Albertyn and Goldblatt*, (note 14) at 253.

⁵² See note 10.

⁵³ See note 10.

⁵⁴ *Hugo* para 41, quoted approvingly in *Harksen* at para 50.

⁵⁵ See note 10.

⁵⁶ See *National Coalition (No 1)* (note 10) at para 17.

⁵⁷ At para 51.

a non-exhaustive list.⁵⁸ Although the Court in *Harksen* specifically stated that it was the cumulative effect of the three factors which was to be considered in deciding the unfairness question⁵⁹, the majority judgment, at the same time, appeared to give prominence to the protection of human dignity, quoting the above dictum from *Hugo* with approval.⁶⁰ The enunciation of the *Harksen* test has been regretted by some commentators, as representing a shift in focus from an overtly remedial equality provision which is designed and operates so as to cure historical, systemic disadvantage, epitomised, it is argued, by *Brink*, to one which is less purposefully conceived. The perceived privileging of the concept of dignity – at the expense of the “value of equality” – is seen as eviscerating the essence of the equality clause: dignity being a liberal, individualised conception of the right to equality.⁶¹ That a conception of equality which focuses exclusively on redressing the wrongs of the past was central to the jurisprudence of the Court in the earlier cases is not perhaps so self-evident from the judgments. Although the statement of O’Regan J in *Brink* appears to accept group-based historical disadvantage as the sole criterion relevant to the assessment of discrimination, it is equally plausible that this approach was merely reflective of the facts of the case. There is nothing in her judgment which precludes other criteria from being identified as relevant. And such other criteria do emerge in subsequent cases. It is also arguable that the impression that the notion of human dignity has displaced historical disadvantage as the central concern of the equality clause is not borne out by the application of the *Harksen* test to the facts of the cases, in spite of the prominence given to human dignity in some passages.⁶²

Although disadvantage is now clearly established as a factor in determining unfairness, there has been little discussion of the meaning of disadvantage itself. In the *Harksen* test, the question to be asked is simply whether the complainant “had suffered in the past from patterns of disadvantage”.⁶³ But clearly, different types of disadvantage arise in the cases. In *Brink*, for example, the Constitutional Court held that certain provisions of the Insurance Act⁶⁴ discriminated against women. The aim of the provisions in question was to protect the interests of creditors of insolvent estates, but it did so by depriving married women in some circumstances of some or all of

⁵⁸ Goldstone J stated in clear terms in *Harksen* that the three considerations do not constitute a closed list and that others may emerge in the course of the development of the Court’s equality jurisprudence (at para 51).

⁵⁹ At para 51.

⁶⁰ At para 50. See also *Walker* (note 10) which succeeded on the basis that the policy operated by the local authority affected the applicants in a manner which was comparably serious to an infringement of dignity.

⁶¹ See Albertyn and Goldblatt (note 14) at 258 and Davis, “Equality: The majesty of Legoland Jurisprudence” (1999) 116 *South African Law Journal* 394 at 404.

⁶² It is arguable that the deciding factor in three of the cases (*Hugo*, *Prinsloo and Harksen*) was the aim of the provision. In *National Coalition (No1)* and (*No2*) and *Larbi-Odam* historical disadvantage and impairment of dignity support and reinforce each other. In *Walker* impairment of human dignity is given prominence, the case succeeding on the basis that the discriminatory measure had an impact which was comparably serious to the infringement of human dignity.

⁶³ *Harksen*, para 50-1, per Goldstone J.

⁶⁴ Act 27 of 1943.

the benefits of life insurance policies ceded to them or made in their favour by their husbands. The disadvantage at issue here was monetary or material. In the course of giving judgment in *Brink*, O'Regan J also commented on the disadvantages suffered by black people in the past as a result of racial discrimination:

“Our history is of particular relevance to the concept of equality. The policy of apartheid, in law and in fact, systematically discriminated against black people in all aspects of social life. Black people were prevented from becoming owners of property or even residing in areas classified as ‘white’ which constituted nearly 90% of the landmass of South Africa; senior jobs and access to established schools and universities were denied to them; civic amenities including transport systems, public parks, libraries and many shops were also closed to black people.”⁶⁵

Clearly, disadvantage does not relate to material matters only, but may be suffered in relation to education, job opportunities and access to public amenities. In *National Coalition for Gay and Lesbian Equality and Another v Minister of Justice and Others (National Coalition No 1)*, which concerned the constitutionality of a number of provisions criminalizing gay sex, different types of disadvantage experienced by gay people as a result of the criminal provisions in question were identified:

“Even when these provisions are not enforced, they reduce gay men . . . to what one author has referred to as ‘unapprehended felons’, thus entrenching stigma and encouraging discrimination in employment and insurance and in judicial decisions about custody and other matters bearing on orientation.”⁶⁶

The Court also emphasised the psychological harm suffered by gay people as a result of the impugned provisions, noting that the existence of criminal sanctions against homosexual acts reinforced prejudice against gays and generated feelings of anxiety and guilt which could lead to depression or have other serious consequences.⁶⁷ Clearly, then, the term disadvantage as used by the Constitutional Court has a wide and inclusive meaning, providing the Court with the opportunity to consider a range of circumstances which impact on applicants, which in turn provides scope for broad application of the non-discrimination provision.

Investigating Disadvantage

Harksen represents an attempt by the Constitutional Court to state a principled and comprehensive test consistent with the overall purpose of the equality provision. The goal is explicitly and emphatically to reflect a substantive approach to equality. The test itself is, of course, not

⁶⁵ At para 40.

⁶⁶ At para 23 quoting Cameron “Sexual orientation and the Constitution: A test Case for Human Rights” (1993) 110 *South African Law Journal* 450 at 455.

⁶⁷ At para 23 citing the decision of *Norris v Republic of Ireland* (1991) 13 EHRR 186 at 192 para 21.

unproblematic: the identification of that range of factors which are relevant to claims under the equality provision renders it vulnerable to criticism as both over-inclusive and imprecise. Presumably, disadvantage is relevant to the test formulated in *Harksen* because some weight is to be given to the fact, thereby enabling the court to give legal effect to that finding. It is worth recalling that the test is formulated so as to consider first, whether the claimant has suffered in the past from patterns of disadvantage, second, the aim of the provision and third, with due regard to the other factors, whether the discrimination has affected the rights or interests of the claimant or has led to an impairment of fundamental dignity (or some comparably serious impairment). Suppose that the test is applied and there is a finding that a claimant has suffered in the past from patterns of disadvantage. What will the consequences of such a finding be? What if the patterns of disadvantage are subject to other remedial measures, such as affirmative action: are these relevant to a claim under section 9(3)? Is there a distinction to be drawn between two claimants under section 9(3): one who has suffered in the past from patterns of disadvantage, and one who has not, where their claims are otherwise in all relevant respects similar? Although the Constitutional Court has begun to grapple with some of these questions, a significant number of their responses remain controversial.

Hugo is illustrative. At issue in the case was a Presidential Act, which granted remission of sentence to women prisoners with children under the age of 12 years. According to the applicant, this discriminated against men with children under 12. All the members of the court agreed that the Act discriminated on a specified ground, namely sex. There was also general agreement regarding the test for unfairness, but two of the members of the Court entered strong dissents based on the application of the test to the facts of the case.⁶⁸ The majority of the court, *per* Goldstone J, held that men were not a historically disadvantaged group, that the aim of the provision was to benefit women who had been disadvantaged in the past, and that the action did not impair the fundamental dignity of the applicants. But these findings followed an almost mechanical application of the test, in which disadvantage was often repeated, mantra-like, without any real investigation of the role and consequences of the particular disadvantage in question, leading to the conclusion that the discrimination imposed by the Presidential Act was not unfair. Kriegler J was highly critical of this verdict, particularly the prominence given in the majority judgment to the conclusion that the act would benefit women. In his view, the decision to pardon women and not men was based explicitly on the assumption that women were the primary caregivers, an assumption which had in the past been the cause of women's inequality. Thus, instead of reversing patterns of disadvantage, the Act had the effect of perpetuating a particular view of women which had in the past operated to their detriment. Kriegler J accepted that there may be circumstances in which Acts which in fact perpetuated disadvantage could be used in order to rebut a presumption of unfairness, but he was emphatic that this was only possible if the advantages of perpetuating a particular stereotypical view clearly and sufficiently outweighed the disadvantages. On the facts of the case he held that the benefits to a relatively small number of

⁶⁸ Kriegler J and Mokgoro J. Didcott J dissented for reasons not connected with the application of the non-discrimination clause itself.

women released from prison were far outweighed by the detriment to both men and women in the perpetuation of gender stereotypes associated with childcare. In her separate dissenting judgment, Mokgoro J, too, highlighted the negative consequences of perpetuating the view that it was women who bore the primary responsibility for childcare. However, her judgment emphasised the impact of this view on men. Treating men as less able parents she said, constituted “an infringement upon their equality and dignity”⁶⁹, which supported the conclusion that the action of the President constituted unfair discrimination.⁷⁰ It is this bolder vision of the role of the equality clause and an engagement with the broader impact of discrimination which distinguishes the judgments of Kriegler and Mokgoro JJ. By contrast, the judgment of Goldstone J is notable for its failure to undertake more than a superficial analysis of the meaning of equality and its largely uncritical acceptance that action which benefited women, any women, must necessarily be fair. Although O’Regan J agreed with the conclusions reached by Goldstone J, she attempted to respond to the objections raised by Kriegler J and Mokgoro JJ. In her view, the Presidential Act benefited women prisoners while causing no significant harm to other women, and it did not harm male prisoners either, since it did not deprive them permanently of the benefits of parenthood. However, in focusing only on the immediate benefits and disadvantages to the parties involved, her judgment served to obscure rather than enlighten the question of the precise role of disadvantage in determining unfairness or the wider role of the equality clause.⁷¹

The precise legal effect of disadvantage is furthermore obscured by terminological inconsistencies. Although it is disadvantage which is identified in the *Harksen* test as being relevant, disadvantage is often used interchangeably with, or linked to, the concept of “vulnerability.” For example, in *Larbi-Odam*⁷² Mokgoro J, giving the unanimous verdict of the Court, concluded that discrimination against non-citizens was unfair. But, in applying the first part of the unfairness test which relates to historical disadvantage, the only factor specifically mentioned was that non-citizens were a vulnerable group.⁷³ In *National Coalition (No 1)*, Ackermann J noted:

“The impact of discrimination on gays and lesbians is rendered more serious and their vulnerability increased by the fact that they are a political minority not able on their own to use political power to secure favourable legislation for themselves.”⁷⁴

Since both *Larbi-Odam* and *National Coalition (No 1)* were concerned with groups who had been disadvantaged, it is not necessarily significant that the court also referred to those groups as vulnerable. But in *Walker*, in his majority judgment in the case, Langa DP, said:

⁶⁹ At para 92.

⁷⁰ Mokgoro J however ultimately concluded that the act was justified in terms of section 33 of the Constitution, (now s 36, see note 22).

⁷¹ A similar difference in approach is evident in the majority and minority judgments in *Harksen*.

⁷² See note 10.

⁷³ At para 23.

⁷⁴ At para 25.

“The respondent does however belong to a racial minority which could, in a political sense, be regarded as vulnerable. It is precisely individuals who are members of such minorities who are vulnerable to discriminatory treatment and who, in a very special sense, must look to the Bill of Rights for protection.”⁷⁵

This was said in almost the same breath as noting that the applicant belonged to a group which had not been economically disadvantaged in the past.⁷⁶ The question then arises: what exactly is the relationship between disadvantage and vulnerability? This question has been partially answered in *National Coalition (No 2)*, by Ackermann J who said:

“This Court has recognised that ‘[t]he more vulnerable the group adversely affected by the discrimination, the more likely the discrimination will be held to be unfair.’⁷⁷ Vulnerability in turn depends to a very significant extent on past patterns of disadvantage, stereotyping and the like. This is why an enquiry into past disadvantage is so important.”⁷⁸

Although, according to the Court, then, disadvantage is an indicator of vulnerability which in turn weighs in favour of a finding of unfairness, it is also clear from the passage that vulnerability is not entirely dependent on past disadvantage and could be proved in other ways, such as was the case in *Walker*. The judgment of Ackermann J in *National Coalition (No2)* also for the first time makes explicit the relevance of disadvantage to discrimination, quoting at length from the Canadian case of *Law v Canada (Minister of Employment and Immigration)*⁷⁹:

“[P]robably the most compelling factor favouring a conclusion that differential treatment imposed by legislation is truly discriminatory will be, where it exists, pre-existing disadvantage, vulnerability, stereotyping or prejudice experienced by the individual or group. . . These factors are relevant because, to the extent that the claimant is already subject to unfair circumstances or treatment in society by virtue of personal characteristics or circumstances, persons like him or her have often not been given equal concern, respect and consideration. It is logical to conclude that, in most cases, further differential treatment will contribute to the perpetuation or promotion of the unfair social character-ization, and will have a more severe impact on them, since they are already vulnerable.”⁸⁰

Accordingly, disadvantage is relevant because without an assessment of pre-existing disadvantage it is not possible for the Court to determine the full impact on the applicant of a measure subject to challenge. Reference to

⁷⁵ At para 48.

⁷⁶ At para 47.

⁷⁷ O’Regan J in *Hugo* para 112.

⁷⁸ At para 44.

⁷⁹ (1999) 170 DLR (4th) 1.

⁸⁰ *National Coalition (No 2)* (note 10) at para 44.

“equal concern, respect and consideration” in the quoted passage, however, also raises the vexed question of the relationship between disadvantage and impairment of dignity.⁸¹ That there may be a close relationship, or even an overlap, between disadvantage and infringement of dignity is also clearly demonstrated in *National Coalition for Gay and Lesbian Equality and Another v Minister of Justice and others (National Coalition No 1)*.⁸² Consider the following passage from the Canadian case of *Vriend v Alberta*⁸³ quoted with approval by Ackermann J :

“Fear of discrimination will logically lead to concealment of true identity and this must be harmful to personal confidence and self-esteem. Compounding that effect is the implicit message conveyed by the exclusion, that gays and lesbians, unlike other individuals, are not worthy of protection. This is clearly an example of a distinction which demeans the individual and strengthens and perpetrates [*sic*] the view that gays and lesbians are less worthy of protection as individuals in . . . society. The potential harm to the dignity and perceived worth of gay and lesbian individuals constitutes a particularly cruel form of discrimination.”⁸⁴

However, although the Constitutional Court has frequently cited Canadian dicta⁸⁵ which emphasise the interrelationship of the concepts of disadvantage and dignity, it has not properly analysed the connection between the two concepts and their separate roles in the *Harksen* test.

National Coalition (No2) has therefore begun to clarify the role and legal effect of disadvantage in determining unfairness, but the discussion of these issues in the case is very brief, and further questions such as what weight is to be given to disadvantage in the *Harksen* test and the exact nature of the relationship between disadvantage and impairment of dignity, as separate elements of the test, remain.

Another question, not considered in *National Coalition (No 2)*, is whether the *Harksen* test is to be applied to the individual applicant in any given case or the applicant as representative of particular groups. The test itself is formulated in terms which indicate that it is patterns of disadvantage suffered by the individual claimant which are relevant. But if that is the actual test to be applied, then women such as Mrs Harksen, who come from a privileged position in South African society, would arguably never succeed in any case. That this cannot be a correct understanding of the test is perhaps borne out by the minority judgment, which takes detailed account of the position of women in society rather than Mrs Harksen herself. It must therefore be

⁸¹ This issue has been hotly debated by South African commentators. See for example Albertyn and Goldblatt, (note 14); Davis, (note 61) and Fagan, “Dignity and Unfair Discrimination: A Value Misplaced and a Right Misunderstood” (1998) 14 *South African Journal on Human Rights* 220. For a general consideration of dignity and human rights see Feldman, “Human Dignity as a Legal Value - Part I” [1999] *PL* 682. (Part II is in Spring 2000 *PL* 61).

⁸² See note 10.

⁸³ (1998) 156 DLR (4th) 385.

⁸⁴ At para 69.

⁸⁵ See also *Hugo* at para 41 and *Prinsloo* at para 32.

asked whether, or when, an individual applicant ought to be able to rely on the non-discrimination provision merely because he or she belongs to a disadvantaged group, yet, as an individual has not suffered in the past. There is a clear argument in favour of giving prominence to group disadvantage if the measure at issue in fact or potentially affects the whole group, such as in the case of legislation. But can a distinction be made between such provisions and measures which affect only a particular individual – such as sacking or deprivation of property – who may not have been historically disadvantaged? It is submitted that such a distinction would be very difficult to draw in practice. It is extremely problematic to argue that a particular measure has an impact on a particular individual only, since it is conceivable that any case which is decided on this basis, itself, could have an impact on the way in which the group is treated. Perhaps the use of more than one factor to determine unfairness (especially impairment of dignity) provides the court with a more sophisticated means of distinguishing the deserving from the undeserving. But if that is so, it is important that this be squarely faced by the Constitutional Court and that elaboration or extension of the *Harksen* test be considered in order to deal with a matter which is potentially complex and divisive.

There is no doubt that the inclusion of the disadvantage element of the test will need to be elaborated and refined by the court. Until the recent case of *National Coalition (No 2)*, there had been little real justification for the inclusion of patterns of disadvantage in the unfairness test. Given the different judgments, it seemed that the phrase meant different things to different judges. Although it is understandable that the Court may prefer to move slowly, developing and clarifying the principles on a case by case basis, it is surely incumbent on the Court to do more than expound a factor (which is admittedly highly populist) without undertaking a rigorous examination of its place in the jurisprudence. In the light of the paucity of cases in which the role of disadvantage has been fully examined and explained, it is perhaps useful to consider another jurisdiction which has also attempted to avoid the pitfalls of the formal Aristotelian test in favour of a test for discrimination which looks to disadvantaged status.

A Canadian Comparison

The status of disadvantaged serves a functional – if somewhat uncertain – role in Canadian case law. At issue in the case of *Andrews v Law Society of British Columbia*⁸⁶ was the fact that the appellant could, in all respects, comply with the requirements to practise law in the province of British Columbia, but he was denied entry to the bar because he lacked Canadian citizenship. The court in that case unanimously rejected the Aristotelian test as “seriously deficient” as “the fixed rule or formula for the resolution of equality decisions”.⁸⁷ In its favour the court adopted a test which looks to whether a person is a member of a persistently disadvantaged group: if a person of that group can show that characteristics either personal to the individual, or arising out of being a member of the group, continue or worsen the disadvantage, then the distinction is discriminatory whether intentional or not.

⁸⁶ 1 SCR 143 (1989).

⁸⁷ At 168, per McIntyre J.

The “disadvantaged” test was later applied in a pregnancy case. In *Brooks v Canada Safeway Ltd*⁸⁸ benefit provisions in legislation gave disfavoured treatment to pregnant workers, not shared by men or women who were not pregnant. The Court stated that “[it] is difficult to conceive that distinctions or discriminations based upon pregnancy could ever be regarded as other than discrimination based upon sex”.⁸⁹ In this case, the court found it unnecessary to find a male comparator: the disadvantage pregnant women suffer comes about because of their condition; their difference. The court located the discrimination in the disadvantage of having to bear the costs of pregnancy. As the Court stated, “those who bear children and benefit society as a whole thereby should not be economically or socially disadvantaged. . . . It is unfair to impose all the costs of pregnancy upon one half of the population”.⁹⁰ Once this was decided, it was impossible not to find that differential treatment on the basis of pregnancy was anything but discrimination on the basis of sex.

This test was subsequently applied to find that sexual harassment was without doubt sex discrimination, when the Supreme Court unanimously overturned a lower court’s decision that sexual harassment did not in fact constitute discrimination based on sex.⁹¹ The reasons for the decision were based on the fact that sexual harassment has a negative, differential impact on women in terms of gender hierarchy in the labour force, and that it is inherently an “abuse of both economic and sexual power.”⁹² The Court compared sexual harassment with racial harassment, stating that sexual harassment was as much a barrier to sexual equality as racial harassment was to racial equality.

Brooks represents a re-evaluation of the concept of gender equality. The Supreme Court of Canada favoured a contextual approach over the detached objectivity of the Aristotelian model, so that the concept of equality is given content which is based on the systemic disadvantages in society. In other words, the court is obliged to view the claimant’s social, political and legal reality. This test does not rule out a male comparator where it is relevant because the remedies require identical treatment as between women and men: but where the issues are female-specific, the comparator is abandoned in favour of the question as to whether the law increases the persistent disadvantaged status of women. The effect of a shift from an individual (whether hypothetical or real) male comparator to the identification of a disadvantaged group enables the court to consider systemic discrimination.

The court’s claim that it has abandoned the “similarly situated” test in *Andrews*, in favour of the new “disadvantaged” test subsequently applied in *Brooks*, is somewhat overstated. The court in *Andrews* did not deny that equality is an inescapably “comparative concept”⁹³. Similarly, there cannot be disadvantage without advantage. The test is radical because the concept of proportionality is replaced with a presumption that where those who are different because they are socially disadvantaged are also affected by laws – based on personal characteristics of the individual or group – which

⁸⁸ 1 SCR 1219 (1992).

⁸⁹ At 1244.

⁹⁰ At 1243.

⁹¹ *Janzen and Govereau v Platy Services* [1989] 1 SCR 1252 .

⁹² *Janzen* at 1284.

⁹³ *Andrews* (note 86) at 164, per McIntyre J.

perpetuate that status then they are discriminated against. The result is that discrimination is recharacterised in terms of disadvantage to already disadvantaged persons.

The court clearly conceived of equality and discrimination as distinct concepts which, although they are related, operate independently of each other. This view has been sharply rebuked. As Beatty has remarked:

“Discrimination relies on the concept of relevance to test the adequacy of the reasons that are advanced to justify distinctions and classifications when they are challenged in exactly the same way as the formal equality rule. Relevance is as central to the proscription against discrimination as it is to the principle of equality. In fact, discrimination and equality are virtually identical concepts. They form a kind of binary relationship, a pair of opposites, like two sides of the same coin. Equality and discrimination are words that describe two possible (and contradictory) conclusions that can be drawn after an impugned classification is tested for the extent to which it is relevant to the objectives it is expected to promote.”⁹⁴

The court’s attempt to reformulate the content of equality norms in *Andrews* was based in part on a literal (mis)understanding and application of the equality principle. There, the Court stated, that the application of the equality laws could justify furthering atrocities such as those under the Nuremburg laws. But that would only be achieved if one ignores the way in which the equality principle has been given effect as legal norm. In a review of the *Andrews* decision, Gold⁹⁵ has pointed out that the core idea of formal equality today is that legislative distinctions must be relevant to the purposes of law. The principle of treating likes alike is violated if a law distinguishes between two classes which are similarly situated with respect to the purpose of that law. When this approach is coupled with the limits which constitutional principles place on the legitimacy of legislative purpose, the formal notion of equality is saved from the court’s conclusion that it could legitimate atrocities such as the Nuremburg laws.

Gold’s criticisms might be answered, in part, by observing that he somewhat misstates the modifications of the Aristotelian approach in law. The principle of equality may be violated if a law distinguishes between two classes which are similarly situated with respect to the purpose of that law, but the law is often nevertheless saved because of the constitutional principles which limit the absolute character of human rights norms. Moreover the criticism made by Gold and Beatty are valid in so far as one is able to clearly identify the distinction by which the class is defined. But the virtue of the disadvantage test from a woman’s point of view is that it locates the remedy precisely in the difference and disadvantage which has been ignored under the relevance aspect of the Aristotelian approach in many equality cases. The disadvantage test escapes the need to first decide

⁹⁴ Beatty, “The Canadian Concept of Equality” (1995) *University of Toronto Law Journal* 349 at 361.

⁹⁵ Gold, “Comment: *Andrews v Law Society of British Columbia*” (1989) 34 *McGill Law Journal* 1063.

whether distinctions based on pregnancy, sexual harassment, reproduction, and others are in fact based on “sex”, because the starting point, when one is dealing with women as a disadvantaged group, is the difference (plus disadvantage) itself, and once that is identified, the fact that it is based on “sex” and is therefore a prohibited ground of discrimination, follows.

The Canadian disadvantaged test has numerous difficulties, some of which have been articulated here, but these difficulties, and the implications of the test itself, have not yet been fully explored by the courts. Instead, in some recent cases, the Court has retreated from the “disadvantaged” test. *Egan v Canada*⁹⁶, *Miron v Trudel*⁹⁷ and *Thibideau v Canada*⁹⁸ were decided together in 1995. The majority upheld a definition of equality which is restrictive and literal, and which in the first two cases seems to have led to error. In *Egan* the court upheld a section of the Old Age Security Act which restricted payment of spousal allowance for elderly people in need to married couples of the opposite sex. In *Miron*, “common law” spouses barely persuaded a majority that Ontario’s Standard Automobile Policy discriminated unfairly against them. In both these cases, the majority tested the relevance of the classification of “married” couples literally, without reference to the overarching objectives of the legislation. In other words, they used the wrong standard of comparison, ignoring the “similarly situated” test’s requirement that the purpose of the legislation be the standard by which the classes are evaluated. Even applying the formal equality test used before *Andrews*⁹⁹, they should have looked to how the classification served the broader purposes of the law at issue, the alleviation of the situation of serious economic need. On this application, these two cases probably should have succeeded. But in these cases, it was only L’Heureux-Dube who attempted to make the adverse impact and disadvantage effected by the law the focus of the court’s review.

However, in a recent judgment, the Canadian Supreme Court has attempted to further clarify the concept of disadvantage in relation to a discrimination claim, and has arguably cured some of the errors revealed in the 1995 trilogy. *M v H*¹⁰⁰ concerned a challenge to section 29 of the Family Law Act of Ontario, which limited spousal support applications to same sex couples, but extended support to couples who were not married to each other but who were in a relationship of some permanence. The applicants in the case were a lesbian couple who separated after some 18 years together. The Supreme Court struck down section 29. In doing so the court restated its understanding, enunciated in *Law*, that probably the most compelling factor favouring a conclusion that differential treatment imposed by legislation is truly discriminatory will be, where it exists, pre-existing disadvantage, vulnerability, stereotyping or prejudice experienced by the individual or group.

The Court clearly conceived of disadvantaged status as an impairment of dignity, which is itself exacerbated in those cases where discrimination is found

⁹⁶ [1995] 2 SCR 513.

⁹⁷ 124 DLR (4th) 693.

⁹⁸ [1995] 2 SCR 627.

⁹⁹ Above note 86.

¹⁰⁰ (1999) 171 DLR (4th) 577.

to have been imposed by legislative measures. In the course of the judgment the court took the opportunity to reiterate the contextual nature of the inquiry in an equality claim, and to clarify the legal effect of ameliorative legislation. The single dissenting judge had observed that the Family Law Act was itself ameliorative legislation, aimed at women in opposite sex relationships, and thereafter determined that women in same sex relationships were not similarly disadvantaged. The majority disagreed. They stated the principle that any ameliorative legislation which accords with the purpose of section 15 of the Charter and which excludes a member of an advantaged group, will likely not violate the human dignity of the more advantaged group where their exclusion corresponds with the greater need or different circumstances of the disadvantaged group targeted.¹⁰¹ Furthermore, the court reiterated that the disadvantage of the claimant does not preclude a claim by those not disadvantaged. The first question to be asked under section 15 is whether the provision draws a formal distinction between the claimant and others on the basis of one or more personal characteristics or whether the provision fails to take into account the claimant's already disadvantaged status.

CONCLUSION

The Canadian jurisprudence has challenged the Aristotelian relationship between equality and discrimination by rescuing the remedial aspect of the anti-discrimination law from the unsatisfactory comparator required by the legal application of the Aristotelian equality ideal. And it has done so by undertaking a legal analysis of the concept of disadvantaged status. In South Africa, the Harksen test has the potential to avoid the pitfalls which the Canadian approach encountered over the years. The South African Court has begun to undertake an inquiry as to how disadvantage is legally effective, and has looked in part to the Canadian cases for some guidance. But it is suggested that any wholesale borrowing of the Canadian approach would be unfortunate. The conceptual foundations for equality and discrimination law, the relationship between disadvantage and dignity, and the remedial effect of disadvantage are only a few of the issues which need consideration within the South African context and under the Harksen test itself. The re-evaluation of equality and discrimination undertaken in both Canada and South Africa is in its infancy, and the South African experience should no doubt unpack the conceptual foundations further.

¹⁰¹ At para 71.

**OF LAME DUCKS, BLACK SHEEP AND FAMILY
BONDING**

**ONE CENTURY OF THE WILL-MAKING DUTIES OF A
COMMONWEALTH FATHER**

*Sheena Grattan, Senior Lecturer in Law, Queen's University of
Belfast**

INTRODUCTION

From the perspective of the succession lawyer there can have been few more significant statutory innovations throughout the Commonwealth during the century which has just ended than the “flexible restraint on testamentary freedom”, the phrase first coined by Professor Laufer as a generic term for the type of legislation which will no doubt be most familiar to the present readership as the Inheritance (Provision for Family and Dependants) (Northern Ireland) Order 1979, and the Inheritance (Provision for Family and Dependants) Act 1975, its virtually identical English counterpart. Indeed, such family protection legislation, to introduce another accepted generic label (the one which will be employed throughout this paper), is as old as that century itself, having been pioneered by New Zealand, then the “social laboratory of the world”,¹ in the form of the Testator’s Family Maintenance Act 1900 at its very dawn. Within forty years legislation adapted to local variations and rejoicing in a wide variety of names (for example, “dependant’s relief”, “family provision” and “will variation” legislation), but broadly adopting the seminal structure of this New Zealand prototype had been introduced to Australia and Canada,² and, following a prolonged parliamentary struggle, to England.³ By the latter half of the century similar

* The author would like to thank Professor Gareth Miller, University of East Anglia, for the comments he made on the draft of this article.

¹ During the twenty-one years in which the New Zealand Liberal Party was in power (1891-1912) New Zealand initiated a programme of legislative reform, the creativity and extent of which far belied the country’s size and population. Other legislation enacted included the Factories Act 1891, the Employer’s Liability Act 1891, the Industrial Conciliation and Arbitration Act 1894 and the Old Age Pensions Act 1898.

² In Australia: Victoria – Widows and Young Children Maintenance Act 1906 (since replaced by Victoria Administration and Probate Acts 1928 and 1958); Tasmania – Testator’s Family Maintenance Act 1912; Queensland – Testator’s Family Maintenance Act 1914 (since replaced by Part IV of the Succession Act 1981); New South Wales – Testator’s Family Maintenance and Guardianship of Infants Act 1916 (since replaced by the Family Provision Act 1982); South Australia – Testator’s Family Maintenance Act 1918 (since replaced by Inheritance (Family Provision Act 1972); Western Australia – Testator’s Family Maintenance Act 1939 (since replaced by the Inheritance (Family and Dependants) Act 1972); ACT – Part VIII of the Administration and Probate Ordinance 1929; and Northern Territories – Testator’s Family Maintenance Ordinance 1929.

³ The Inheritance (Family Provision) Act 1938, since replaced by the Inheritance (Provision for Family and Dependants) Act 1975. For a good account of the

enactments had spread to yet more of the English speaking world, including both Irish jurisdictions,⁴ and beyond.⁵

As the century closed, family protection legislation, “succession by arbitration”, had unquestionably established itself as a third “pillar” of the traditionally bi-partite succession edifice (that is, the rules of testate and of intestate succession); a far cry from the observation once made by Lord Reid, when interpreting the provisions of an Act of 1882, that it was “inconceivable” that Parliament should have put the whole of a man’s capital at the discretion of a judge!⁶ It is thus rather surprising that family protection legislation, which has done exactly that, has provoked relatively little academic discussion or debate – at least in this part of the world.⁷ It would be unfortunate if the centenary of this genus of legislation were to pass without some assessment being made of the impact which it has made. An all-embracing review is beyond the scope of a single article, so the focus of this paper has been limited to one particularly influential strand of Commonwealth jurisprudence, namely the analysis of the family protection jurisdiction in terms of the breach of a moral duty owed to an applicant by the testator.

THE MORAL DUTY TEST CONCEIVED

“It is the duty of the Court, so far as possible, to place itself in all respects in the position of the testator, and to consider whether or not, having regard to all existing facts and surrounding circumstances, the testator has been guilty of a manifest *breach of that moral duty which a just, but not a loving, husband or father owes towards his wife or towards his children*, as the case may be”.⁸

When Edwards J of the New Zealand Court of Appeal first pronounced these now familiar words back in 1910 in the course of his judgment in *Allardice v Allardice*, the learned judge probably had little reason to suspect that they would become what is arguably the most oft-quoted judicial dictum in the history of family succession law, or that he had just spawned what Laufer would later christen “the reasonable man of the law of domestic relations”.⁹

parliamentary background, see *Tyler’s Family Provision* (third ed, RD Oughton, 1997).

⁴ In Northern Ireland the Inheritance (Family Provision) Act (NI) 1960, since repealed and replaced by the Inheritance (Provision for Family and Dependents)(NI) Order 1979. In the Republic s117 of the Succession Act 1965. This latter provision extends only to the testator’s children. A parallel fixed share system protects the spouse (s 111 of the 1965 Act).

⁵ *Eg* Israel – chapter 4 of the Succession Law 1965.

⁶ Cited in S Cretney, “Discretion or Whim, Freedom of Choice or Caprice?” (1986) 6 *OJLS* 299.

⁷ A comment which is valid of succession law in general in the United Kingdom. See, *eg*, the comments made by Professor Miller in his preface to *The Machinery of Succession*, (2nd ed, 1996). The same has not been true of succession issues in other parts of the Commonwealth, such as New Zealand.

⁸ *Per* Edwards J in *Re Allardice; Allardice v Allardice* (1910) 29 NZLR 959 at 972-973.

⁹ J Laufer, “Flexible Restraints on Testamentary Freedom” (1955) 69 *Harvard Law Review* 277 at 295.

Since then a not insignificant number of the world's judiciary have been assiduously rectifying what they have deemed to be breaches of moral duty by will-makers from many varied social and ethnic backgrounds, as Edwards J's test evolved into the benchmark for judicial intervention in the discretionary family protection systems of Australasia, Canada and the Republic of Ireland. Amongst these the Republic of Ireland stands unique - to the author's knowledge the only jurisdiction in the world to enshrine Edwards J's hypothetical creature in statute.¹⁰ Elsewhere, however, the absence of statutory recognition has not thwarted the ambitions of this juristic person; he, and his morality, have succeeded in influencing the concept of testamentary freedom far beyond anything which a mere "judicial gloss"¹¹ might have aspired to. By the 1970's and the 1980's it seemed that the years might be taking their toll as the judicious father and husband came under the microscope of a number of Commonwealth law reform agencies. The prognosis was mixed, but he generally survived to fight another day. As a new century begins, however, it seems that time may finally be running short for this judge-made creation, for if the draft Succession (Adjustment) Act is enacted by the New Zealand Parliament he will be expunged by statute from his homeland, which may possibly herald the slippery slope towards global extinction.

It is thus appropriate that the aim of this article is two-fold. Its first purpose is to explore and assess the impact which *Allardice* has had in shaping family succession within the Commonwealth in the course of the last ninety years. An appraisal of how the decision has affected the full range of family members (for example, widow, widower, and infant children) is beyond the scope of this work, so, for reasons which are explained below, the analysis will be limited to one particular family member, the adult child. Secondly, in light of the proposals for legislative reform made recently by the New Zealand Law Commission it is a timely evaluation of how significant the breach of moral duty test should remain in the twenty-first century. Neither task, however, can be properly undertaken without some understanding of the function(s) which family protection legislation is designed to perform. Before, then, we turn to consider the *Allardice* decision and its legacy more closely, a brief reminder of the history, nature and objectives of this particular legislation may be helpful.

FLEXIBLE RESTRAINTS ON TESTAMENTARY FREEDOM - THE BACKGROUND

New Zealand's Testator's Family Maintenance Act 1900 had purported to provide a novel solution to a problem which was as old as private property

¹⁰ Succession Act 1965, s117, which empowers the testator's children to apply for an award out of their parents' estate. Since 1967 the term "moral duty" has also appeared, but not centrally, in the New Zealand legislation: the Family Protection Amendment Act 1967 amended s 3(2) of the Family Protection Act 1955 by prescribing certain matters which must be considered by the Courts in assessing the testator's "moral duty" to his grandchildren (grandchildren being eligible applicants under this Act).

¹¹ From time to time members of the Commonwealth judiciary have reminded their colleagues that Edwards J's just husband and father is a mere gloss on the statute. A selection of these cases is noted below.

and society itself, namely how best to ensure that weak and vulnerable members of a person's family were afforded adequate economic protection from that person's estate after his death. While legal theorists have disagreed on emphasis few have disputed that one of the functions of the law of succession in a civilised society is the preservation and protection of the family.¹² Indeed, the law of succession has been described as "an attempt to define the family in terms of property."¹³

The precise relationship between inheritance structures and the concepts of both "family" and "property" is obviously highly complex, not least because it is an ever-changing one. For the purposes of this article, however, it can be taken to be broadly the case that prior to New Zealand's pioneering legislation the function of protecting the family was delivered, virtually universally, by means of rigid and inviolable rules¹⁴ whereby both beneficiaries and the extent of their benefit were precisely defined, so that a set portion was guaranteed to persons who could claim to be within a certain legal relationship to the deceased (historically most often his children and rather less frequently his wife).¹⁵

Just over one hundred years ago Professor Ely observed that "at a rash guess nineteen-twentieths of the human beings who have ever lived" have also duly died in the sure and certain knowledge that all their worldly goods would thereafter be allocated strictly in accordance with a prescribed order, totally irrespective of any personal wishes which they might have had.¹⁶ The intervening century has undoubtedly vitiated the arithmetic, but the fundamental principle remains valid, providing a salutary reminder, especially to those from the common law tradition, of the fallacy of regarding the will as the key vehicle of intergenerational wealth transfer. In fact, testamentary freedom was (and remains) rare in simpler societies;¹⁷

¹² See *eg* Ely, "The Inheritance of Property" (1891) 153 *North American Review* 54 and Morton, "The Theory of Inheritance" (1894) 8 *Harvard Law Review* 161.

¹³ T P T Plucknett, *A Concise History of the Common Law*, (5th ed, 1956) 711.

¹⁴ The following discussion proceeds on the basis that testamentary freedom is *contrary* to the family's interests. However, this ignores the fact that any assessment of how the right to will protects, or otherwise, the family can only be made in light of the alternative, the intestacy rules, at that moment in time. Historically, it is undoubtedly the case that there were times when testamentary freedom did actually constitute a protection for certain family members: see Miller, *The Machinery of Succession* (2nd ed, 1996) and Unger, "The Inheritance Act and the Family" (1943) 6 *MLR* 215. For the purposes of this article, however, which is concerned exclusively with the twentieth century, testamentary freedom is generally assumed to be capable of being used against the family's interests.

¹⁵ The movement away from blood relatives in favour of the deceased's spouse is one of the major trends in modern Succession Law – see generally, MA Glendon, *The Transformation of Family Law* (1989).

¹⁶ J Ely, "The Inheritance of Property" (1891) 153 *North American Review* 54 at 54.

¹⁷ See generally, Maine, *Ancient Law* (10th ed, 1927). *Deus solus heredem facere potest, non homo* (it is God, not man, who makes heirs) was a well-known maxim of the old Teutonic Law. Even in England, the home of the common law, the will emerged comparatively late in history. Land could not be willed until the enactment of the Statute of Wills 1540, and even then only to a limited extent (tenants by knight-service still had to leave at least one-third to descend under the inheritance law until the enactment of the Military Tenures Abolition Act 1660).

religious codes such as the Sharia¹⁸ generally prescribed (and continue to prescribe) an intricate system of family inheritance, and to this day Scotland and virtually all continental jurisdictions, which also derive from Roman law, reserve a portion of a property-owner's estate for defined family members, the right to will being limited to that fraction which remains.¹⁹

In contrast, on colonisation in 1840 New Zealand had inherited the English common law²⁰ at the zenith of an "interval of unbridled testamentary licence",²¹ unrestricted testamentary freedom having been one of the fundamental tenets of the eighteenth century *laissez-faire* philosophy espoused by influential English writers such as John Locke and William Blackstone, during an era which has become known as the 'golden age of individualism'. Within decades, however, it had become evident that such unfettered testamentary freedom was a power which could be abused by the New Zealand husband or father at the expense of his wife and young children.²² Perhaps unsurprisingly it had been to the traditional fixed share that New Zealand's law reformers (many of whom had a Scottish

In Ireland the corresponding enactments were the Statute of Wills 1634 and the Tenures Abolition Act (Ir) 1662. Similarly, considerable restraints were placed upon the scope of wills of personalty until the fourteenth century – by virtue of the custom of legitim wives and children were guaranteed certain rights which could not be defeated by will. In effect no more than one half of the personal estate could be disposed of by will: subject to local variations a deceased who was survived by a wife and children was limited to disposing of no more than one-third of his personal estate, the so-called "dead's part". The wife was entitled to one-third and such of the children who had not been advanced in the deceased's lifetime were also entitled to one-third each (the bairn's part). Where the deceased was survived by a wife but no children, her and the dead's share increased to one-half each. In Ireland these restrictions were not removed until the enactment of s10 of the Statute of Distributions 1695.

¹⁸ For further detail, see N J Coulson, *Succession in the Muslim Family* (1971) and Pearl and Menski, *Muslim Family Law*, chapter 11 (3rd ed, 1998).

¹⁹ Various names are given to such systems, eg Legitime (Spain); La Reserve (France); Pflichtteil (Germany); Jus Relicti and legitim (Scotland); Forced heirship (Louisiana). In the Republic of Ireland surviving spouses are protected in this manner by virtue of s 111 of the Succession Act 1965.

²⁰ New Zealand was a ceded colony so English law did not apply automatically. The English Laws Act 1858 affirmed the reception of English Law retrospectively to 1840.

²¹ *Per* Lord Simon in *Schaefer v Schuhmann* (1972) 46 ALJR 82 at 90. The last vestiges of economic protection once afforded a widow had been swept away seven years previously by the Dower Act 1833.

²² See in general Atherton, "New Zealand's Testator's Family Maintenance Act of 1900 – The Stouts, The Women's Movement and Political Compromise" (1990) 7 *Otago LR* 202. This article charts the background to the New Zealand legislation examining, amongst other factors, the role played by the Women's Movement. It has also been suggested that the need to restrict plenary testamentary freedom became an issue in New Zealand, and indeed in Australasia generally, long before it had in England because of the absence of the family settlement in the New World – see *eg* Lee, *Manual of Queensland Succession Law* (2nd ed, 1984). In England such family settlements, which generally defined with precision from birth the exact share of the patrimony which each family member was to receive, provided an effective fetter to the absolute testamentary freedom described above. They were, however, limited to the upper echelons of society.

background) had first looked for a solution to this problem, but, perhaps just as unsurprisingly, such an indiscriminating intrusion into private property rights failed to secure majority support amongst the Liberal parliament-arians of the day.²³

The ingenious political compromise²⁴ between this mandatory fixed share system and absolute freedom of testation was a statutory provision based on discretion – a more flexible, targeted solution which accorded better with the prevailing Liberal ideology which had, by now, been forced to retreat significantly from its erstwhile uncompromising freedom of property ethic to endorse state intervention on moral grounds in order to protect the weaker members of society.²⁵ Even then, however, acceptance was not immediate and the original drafting would have to be refined further to underscore that it was a *moral* basis which was sanctioning this interference with private property rights – this involved the insertion of a “disentitling provision” to clarify that a court could refuse to make an order in favour of any person whose “character or conduct” was undeserving of such help.²⁶ On the third time of asking, however, the discretionary family protection mechanism gained the approval of a majority within the House of Representatives and the Testator’s Family Maintenance Act, a short enactment of only twenty-two lines, received the Royal Assent on 9 October 1900. Its key provision was as follows:

“Should any person die, leaving a will, and without making therein adequate provision for the proper maintenance and support of his or her wife, husband, or children, the Court may at its discretion, on application by or on behalf of the said wife, husband, or children, order that such provision as to the said Court shall seem fit shall be made out of the estate of the said deceased person for such, wife, husband or children: Provided that the Court may attach such conditions to the order made as it shall think fit, or may refuse to make an order in favour of any person whose character or conduct is such as in the opinion of the Court to disentitle him or her to the benefit of an order under this section.”

Such a mechanism was obviously in stark contrast to all existing fixed share approaches. Its claim to novelty was grounded in the innovatory

²³ Sir Robert Stout introduced two fixed share Bills to the New Zealand Parliament, the Limitation of the Powers of Disposition by Will Bills 1896 and 1897. See generally Atherton, *ibid*.

²⁴ The draftsmen responsible for this new formula were undoubtedly influenced by a legislative precedent of recent origin which also protected weaker members of society, s46 of the Native Land Court Act 1894. This provision empowered the Native Land Court to override wills of Maori testators in certain circumstances, a pragmatic solution to the common problem of forged Maori wills being presented in order to assert title to particular lands, in that it circumvented the need to prove fraud, always a difficult evidential hurdle. See further, Atherton *op cit* n 22.

²⁵ It was essentially the aftermath of the Industrial Revolution which exposed the complete inability of classical liberalism to protect the weaker members of society and thus precipitated this change. In England probably the most influential writer to articulate this new ideology was John Stuart Mill.

²⁶ See Atherton, *op cit* n 22.

discretionary concept which it embodied – the legislature had not prescribed a uniform solution to be applied to every estate, but had merely empowered the Court to intervene on application by defined family members if, in the circumstances, it was necessary to vary the deceased’s will because he had not adequately discharged his obligations to his family. Such a discretionary model could claim one very obvious advantage over its mandatory counterpart, namely the delivery of individualised justice. Unlike the mathematical fixed share, it was not forced to treat alike the “deserving and undeserving, rich and poor, old and young, strong and weak, burdened with small children and childless”.²⁷ Equally, however, it would not have needed a crystal ball to foresee that this immense flexibility could also quickly become its greatest weakness.

Legislation which empowers the judiciary to alter a deceased’s testamentary disposition in favour of certain family members or other dependants may be framed with one of at least two different (and potentially conflicting) objectives in mind. Perhaps most obviously, it may be intended to serve a “support” function (that is, it is designed to provide for those in financial need, and possibly only for those whom the deceased had been required to support when he was alive). Alternatively, it may be intended to serve what could broadly be described as a “fair distribution” function (for example, it is designed to reward good conduct, or to provide recompense for past services, or to redress the deceased’s lifetime failings, or, indeed, any combination of these). The former will involve a restrictive approach, connoting terms such as “maintenance”, “income” and “dependence” and will be concerned exclusively with an applicant’s economic need; the latter will involve an expansive approach, connoting terms such as “contribution”, “capital” and “restitution” and will necessarily be concerned with matters beyond the economic status of an applicant.

With the Testator’s Family Maintenance Act the New Zealand legislature granted the widest possible discretion to the Courts, but made no attempt to define the objectives which were to control its exercise, or to provide clear, ordered principles for guidance. Responsibility for determining what constituted “adequate provision for the proper maintenance and support” of an applicant was therefore abrogated entirely to the judiciary. A closer examination of the legislative background of this Act, however, would suggest that the absence of well-defined statutory guidelines was not an oversight, in that the tension between the “support” and “restitutive” functions of the proposed statute was already evident in the debates of the New Zealand Parliament which preceded its enactment. Although a minority of the House of Representatives favoured a broad approach which could take into account the claimant’s station in life, rather than one limited to bare subsistence, (and the occasional Member even favoured a purely “dynastic” approach in which the primary objective was maintaining the patrimony within the family), the majority clearly saw the proposed legislation as an extension of the *inter vivos* support obligations contained in the Destitute Persons Act 1894, existing legislation which empowered the Court to order a living person to maintain certain destitute family members. Thus much of the debate pared down to a straightforward economic argument of “state” versus

²⁷ Atkinson, *Wills* 679 (2nd ed, 1953), cited in Laufer, *op cit* n 9 at page 279.

“estate”, and the support given to the Measure before them by many Members of the House was entirely due to the fact that a testator could no longer foist his disinherited widow and dependent children on the public purse for support. Certainly, Robert McNab (although himself an advocate of the broader approach)²⁸ cast his last Bill firmly in terms of maintenance:

“[this Bill says] before you dispose of your property, first carry out your obligations; first see that you do not leave any person destitute; first see that any person who is at present dependant [*sic*] on you for his or her support and maintenance is not left on the state for support.”²⁹

However, as has been seen, the wording which was finally enacted had a breadth and a generality such that it would be capable of being many different things to different people, including different judges. As one New Zealand commentator has observed:

“The Parliamentary Debates clearly reveal that there was no unanimous view as regards the nature and purpose of this legislation, but it was phrased sufficiently broadly to accommodate different as well as conflicting interpretations. It was the result of a political compromise and left the courts with the task of determining its scope”³⁰

It will be argued in the course of this paper that in the jurisdictions under consideration the Courts have generally struggled to determine the scope of the wide discretion granted to them by family protection statutes, with the result that, at times, the law in this area has been characterised by uncertainty and inconsistency, which should at least provoke the question “have the inherent weaknesses of the flexible restraint on testamentary freedom now started to outweigh its obvious benefits?” First, however, it is necessary to turn the clock back to Dannevirke, New Zealand on 11 March 1908, the date which witnessed the passing of a certain James Allardice Esq, and thus triggered the events which would culminate in the litigation which forms the basis of this paper. It is submitted that it is this decision, more than any other, which has been pivotal in setting the pattern for nearly a century of family protection cases throughout much of the English-speaking world.

ALLARDICE V ALLARDICE

In common with so many fellow testators who have also come to know posthumous fame through estate litigation in its various guises, the late James Allardice had enjoyed a colourful existence. His progression from a simple farmhand, whose only recognised talent was for breaking-in unruly horses, through storekeeping, to hotel ownership and the fortuitous acquisition of property at just the right time ensured that when he died in March 1908 he was able to leave behind a fortune of nearly £ 30,000,

²⁸ The original wording of the Bill introduced by McNab was “due provision for maintenance and support”, but once it had passed its Second Reading, he strategically changed the wording at Committee Stage to “proper” maintenance and support, a word which was to become pivotal in the later interpretation of the Act. See Peart, “Direction of the Family Protection Act 1955” [1994] *NZRLR* 193.

²⁹ *NZPD* (1900) vol 111 cols 503 and 504.

³⁰ *Per* N Peart, *op cit* n 28 at 196.

notwithstanding that his latter years were spent as “a drinking man of loose habits who must have squandered a good deal of his earnings.”³¹ Affairs of the heart had proved no less colourful, and he also left behind a widow (Agnes), a former wife, (the first Mrs Allardice), and a total of thirteen children (seven by Agnes, the first six born out of wedlock, the remaining six by the first Mrs Allardice) ranging in stature from toddler to woman of nearly forty. For five years he had even succeeded in enjoying the company of the two woman simultaneously, but this arrangement was brought to an abrupt end in 1891 when the first Mrs Allardice discovered Agnes, ensconced in a house not far away at her husband’s expense. From this time he lived with Agnes until his death, eventually marrying her in 1905 after a change in the law had enabled the first Mrs Allardice to obtain a divorce. To this large fortune and web of domestic relationships one must add the late Mr Allardice’s will, which benefited only Agnes and her brood, and it should already be obvious that there existed all of the prerequisites for a long, expensive and potentially unsavoury family protection claim. In this regard it was not to disappoint, generating a mass of “very voluminous” evidence contained in thirty-seven affidavits, most of the which Chapman J, the judge at first instance, dismissed as “altogether irrelevant”.³²

In fact all but one of the deceased’s six children by the first Mrs Allardice (three married daughters and two unmarried sons, all “thirty-somethings” by the time of their father’s death)³³ applied for an award out of his estate under section 33 of the Family Protection Act 1908, the recently enacted consolidation of the earlier legislation,³⁴ on the basis that his will did not make “adequate provision” for their “proper maintenance”. None of the five enjoyed a financial position which could have been described as “flourishing”. The three daughters had all been “educated in as good a manner as could be secured in Dannevirke”, primarily through the efforts and exertions of their mother. This had enabled them to secure husbands “of good character” and “of some capability, all of whom [were] able to work”, but none of these men was comfortably off. One daughter, a Mrs Hawkins, was married to a journalist who at the time of the application had just embarked upon a small drapery business. Another, Mrs Hall, had been widowed but was now remarried to a man currently investing in a not too profitable sawmill venture. The third, Mrs Hazeldon, was the least well-off, being married to a clerk whose earnings the judge described as “precarious”.

³¹ *Per* Chapman J at p 960.

³² *Ibid.*

³³ The sixth, a daughter, had “married well” and made no claim. The first Mrs Allardice had no claim because she had been satisfactorily provided for in the divorce settlement.

³⁴ The Testator’s Family Maintenance Act 1900 had been speedily amended by the Testator’s Family Maintenance Act 1906 in order to reverse the majority decision of the Court of Appeal in *Plimmer v Plimmer* [(1906) 9 NZGLR 10] that there was no jurisdiction to award a lump sum under the Act. The decision in *Plimmer* was delivered on 11 July 1906, the amending legislation was passed just over three months later on 19 October. This legislation plus a couple of minor amending statutes were then consolidated in the 1908 legislation. The key substantive provision of the 1908 Act remained the same as that in the 1900 Act, that is, had the deceased made “adequate provision for the proper maintenance and support” of the applicant?

Her brothers appeared to be in a similarly precarious financial situation. They continued to live with their mother on her small thirty acre freehold farm while working as labourers on this land. One had been trained as a saddler at his father's expense, previously embarking upon an unsuccessful business venture in this field on which his father lost the £250 he had put up as a guarantee; the other possessed no particular skill, save that he appeared to have inherited his father's flair for horse-breaking. In short, they were two able-bodied men, whose only property was their personal effects, but neither of whom, to quote Chapman J, seemed to "exhibit any ambition to rise above their present position." The learned judge summed up the problem before him as follows:

"What I really have to consider in this case is whether the Act has any application whatever to able-bodied sons who are capable of supporting themselves in the condition in which they have hitherto lived and who have done so ever since they came to man's estate, or to married daughters who never have been dependent on their father since they were married and whose husbands are capable of supporting them in the future as they have supported them in the past".³⁵

Chapman J duly answered both of his questions in the negative, refusing to make an award in favour of any of the applicants. It had been submitted on behalf of the five children that the deceased had failed to make proper provision for them – more specifically, that the deceased should have attempted to redress the fact that he had failed to educate his sons in a manner more in keeping with his wealth, and that in the case of his daughters he had failed to supplement their slender means.

In rejecting all of the applications, Chapman J emphasised that he was taking "his guidance exclusively from the statute".³⁶ It was evident that the learned judge had a certain sympathy for the three female applicants, but felt that he was not empowered to rectify the injustice which he believed had been done to them – he could have no concern, he stated, with what their "father ought in justice have done beyond the standard [which he] derived from the statute." This "standard" he clearly defined in terms of the relief of destitution, and the three women, while married to men on low wages, could not properly be described as being "in want".³⁷ Significantly, also, they, like their brothers, would have no legal claim to be supported by their father if he were still alive since they were outside the scope of the Destitute Persons Act.

The applicants appealed to the New Zealand Court of Appeal on the ground that the trial judge had exercised his judicial discretion on erroneous principles in that he had failed to put due weight on the requirement that the maintenance be "proper". More specifically, their counsel submitted, Chapman J had not given consideration to the relatively large value of the estate and to the natural obligation which a parent has to make provision for the "proper maintenance" of his children. The Court of Appeal (Stout CJ,

³⁵ At p 964.

³⁶ At p 964.

³⁷ At p 966.

Cooper, Edwards and Williams JJ) unanimously affirmed the judge at first instance with reference to the two able-bodied sons, but reversed the decision in relation to the three married daughters, awarding Mrs Hazeldon £60 per annum and the others each £40 per annum.

Sir Robert Stout, now the Chief Justice, but who during his time in the House of Representatives in the 1890's had been the Member responsible for introducing the two aborted fixed share Bills, delivered the leading judgment, summarising the principles which had been established in the cases decided in the first decade of family protection as follows:

“1, That the Act is something more than a statute to extend the provisions of the Destitute Persons Act; 2, that the Act is not a statute to empower the Court to make a new will for the testator; 3, that the Act allows the Court to alter a testator's disposition of his property only so far as it is necessary to provide for the proper maintenance and support of ‘wife, husband or children’ where adequate provision has not been made for their proper maintenance and support by the will of the testator; 4, that – in the case of a widow, at all events, if not in the case of a widower - the Court will make more ample provision than in the case of children, if the children are physically and mentally able to maintain and support themselves.”³⁸

Both Sir Robert and Cooper J respectively gave examples in the course of their judgments of the type of matters which the Court might take into account when exercising its discretion:

“The Act has laid down no rule, and has left the decision in every case to the discretion of the Court. What, then, has the Court to consider: Firstly, I think, the means of the children. ‘Support’ it has been held – at all events, in the case of a widow – does not mean merely having a supply of food and clothing. It means, it has been held, such kind of maintenance as the widow during the life of her husband has been accustomed to. The matter that should be considered, both as to widow and children, is how she or they have been maintained in the past. A child, for example, that has been living on a father's bounty could not be expected to begin the battle of life without means. A child, however, who had maintained her or himself, and had perhaps accumulated means, might well be expected to be able to fight the battle of life without any extraneous aid. But even in such a case, if the fight is a great struggle, and some aid might help and the means of the testator were great, the Court might, in my opinion, properly give aid. The whole circumstances have to be considered. Even in many cases where the Court comes to a decision that the will is most unjust from a moral point of

³⁸ At p 969.

view, that is not enough to make the Court alter the testator's disposition of his property."³⁹ and

"In my opinion it is impossible to lay down, except in very elastic terms, any general rule in such a case, for many factors have to be considered. The condition of life of the testator and the children, the sex, age and health of the children, whether, if daughters, they are married, or, if sons, they are of sufficient health and strength to earn a reasonable maintenance for themselves or those dependent upon them. . . But. . . I repeat that we have no power to recast the testator's will or to redress inequalities or fancied injustice. . .".⁴⁰

Notwithstanding that both men were careful to draw attention to the fact that they were not empowered to rewrite wills found "unjust from a moral point of view", it is evident that their approach to the exercise of discretion under the 1908 Act was much more subjective than that which had been adopted by Chapman J. The destitution of the applicant, seemingly a *sine qua non* for judicial intervention at first instance, was now one of a variety of matters to be considered relevant, including the wealth of the deceased, the means of the children, and their sex, age, health and marital status. It was Edwards J, however, who crystallised this thinking in the famous words which appeared earlier in this paper:

"It is the duty of the Court, so far as possible, to place itself in all respects in the position of the testator, and to consider whether or not, having regard to all existing facts and surrounding circumstances, the testator has been guilty of a manifest breach of that moral duty which a just, but not a loving, husband or father owes towards his wife or towards his children, as the case may be. If the court finds that the testator has been plainly guilty of a breach of such a moral duty, then it is the duty of the Court to make such an order as appears to be sufficient, but no more than sufficient, to repair it."⁴¹

When their Honours applied these principles to the facts before them they were unanimous that no award should be made in favour of either son, who:

"If they had any push they should, considering their age, have ere this done something for themselves, and to settle money on them now might destroy their energy and weaken their desire to exert themselves".⁴²

In the case of the daughters, however, annuities were awarded, the main factor influencing the Court being the likelihood that in the future their husbands might no longer be able to support them.

The plaintiffs appealed to the Privy Council which approved the decision⁴³ and general approach of the New Zealand Court of Appeal, but in a "cursory

³⁹ *Per* Stout CJ at 969-970.

⁴⁰ *Per* Cooper J at 974-975.

⁴¹ At p 972

⁴² *Per* Stout CJ at 971.

⁴³ [1911] AC 730.

manner”,⁴⁴ with a good part of the short opinion which was delivered by Lord Robson on behalf of the Board given over to observations about the inappropriateness of a body in London interfering with the exercise of discretion by “local courts [which] are well acquainted with all the local conditions as to employment, standard of living and other matters necessary to be borne in mind.”⁴⁵

It should already be evident that it is not the actual decision reached by the New Zealand Court of Appeal on the facts before it which justified the accolade made earlier that *Allardice* is one of the most pivotal family protection cases, for on that level, notwithstanding the awards made in favour of the daughters, it was a relatively narrow interpretation of the relevant statutory provision. The outcome, two young men sent back empty-handed to their mother’s farm, with judicial exhortations about the virtues of self-reliance probably still ringing in their ears, was completely in line with the existing jurisprudence on claims from able-bodied adult children:

“Adult children capable of maintaining themselves may come within the statute of 1900. As to this, I express no opinion. It would, however, I think require a very strong case to justify the Court in making an order under that statute in favour of such a child overriding the will of the testator.”⁴⁶

Rather, the main significance of *Allardice* is found in the gate which Edwards J had firmly opened with his enunciation of the “breach of moral duty” principle, the practical outworkings of which invited the Court to consider a wide range of factors, the ethical as well as the economic. Although the judiciary would tend to tread very warily through this gate for some years, they would eventually succeed in exploring the full extent of the terrain which lay beyond, gradually introducing more and more matters as relevant to the assessment of a testator’s moral duty. It would only be then that the magnitude of the inroads which the decision in *Allardice* had made to the concept of testamentary freedom would become apparent.

Meanwhile, in 1938 the Privy Council enjoyed a further opportunity to consider a Commonwealth family protection case, *Bosch v Perpetual Trustee Co*,⁴⁷ an appeal from New South Wales concerning an application under the Testator’s Family Maintenance and Guardianship of Infants Act 1916.⁴⁸ Their Lordships expressly affirmed the “moral duty” test from *Allardice*, but reformulated it in slightly different terms:

“in every case the Court must place itself in the position of the testator and consider what he ought to have done in all the

⁴⁴ *Tyler’s Family Provision* (3rd ed, RD Oughton, 1997) at p 8. The matter was argued for less than a full day since part of the day was spent hearing *Samaradiwakara v De Sarum* [1911] AC 753.

⁴⁵ *Per* Lord Robson at 734.

⁴⁶ *Per* Edwards J in *Rush v Rush* (1901) 20 NZLR 249. *Rush* was the first reported case on the 1900 Act.

⁴⁷ [1938] AC 463.

⁴⁸ The key provision of this legislation was s 3 which empowered the testator’s spouse and children to apply for an award out of his estate if they had been left “without adequate provision for their proper maintenance, education or advancement in life as the case may be.”

circumstances of the case, treating the testator for that purpose as a wise and just, rather than a fond or foolish, husband or father.”⁴⁹

The judicious husband and father, now also credited with the virtue of wisdom,⁵⁰ and already accepted in Canada,⁵¹ as well as Australasia, had, in the words of one senior member of the New South Wales Bar, been “virtually endowed with perpetual succession”.⁵² And so the “judicial gloss” which had first emanated from the New Zealand Court of Appeal in *Re Allardice*, would continue to dominate Commonwealth family protection law, such that, some forty years later, one New Zealand judge would observe:

“[t]he concept of moral duty, what the wise and just testator would have done, is elementary in the administration of the Family Protection Act in New Zealand. It is too deeply embedded to be open to judicial reconsideration now.”⁵³

We turn next to chart the development of this “moral duty” approach since its inception in 1910, focusing, in particular, on how “what the wise and just testator ought to have done” has altered with the passage of time. For this we leave behind the wise and just husband and concentrate exclusively on how the Australasian and Canadian judiciary have defined the testamentary obligations owed by a just and wise *father* towards his grown-up able-bodied offspring. In one respect, however, it is impossible to leave behind the judicious husband completely. As with any “cake-cutting” exercise, family protection awards made to one particular category of applicant can only be analysed meaningfully in the context of competing claims from other classes of applicant. However, *Allardice* confirmed, and later jurisprudence has entrenched, that a husband’s responsibilities to his wife and infant children are paramount. Many cases which culminate in litigation will not therefore involve adult children competing with the deceased’s spouse (especially if that person is also the applicant’s other parent) or infant children – but with their equally mature siblings, remoter family members, or charities.

The adult child (an eligible claimant under many Commonwealth statutes) has been highlighted because it is he or she who represents the “sharp end”⁵⁴ of the family protection jurisdiction - what Edwards J in *Allardice* described as “the extreme case”.⁵⁵ It is thus in relation to this category of applicant that one might expect any misgivings about the workability of the “moral duty” approach to be most readily confirmed or dismissed. This article is not

⁴⁹ *Per* Lord Romer at p 474.

⁵⁰ The notion of wisdom was first introduced in *Allen v Manchester* [1922] NZLR, a case which was expressly affirmed by the Privy Council in *Bosch*.

⁵¹ The Supreme Court of Canada endorsed the moral duty test in *Walker v McDermott* [1931] DLR 662.

⁵² C McLelland, “Fifty Years of Equity in New South Wales- A Short Survey” (1951) 25 *Australian Law Journal* 344 at 344.

⁵³ *Per* Cook J in *Re Z* [1979] 2 NZLR 495 at 506.

⁵⁴ Green, “The Englishwoman’s Castle - Inheritance and Private Property Today” (1988) 51 *MLR* 187 at 196.

⁵⁵ At p 973. Edwards J was referring to an adult son, but now, as shall be seen, this could be applied equally to an adult able-bodied daughter.

directly concerned with the Irish judiciary's application of the statutory "breach of moral duty" test found in section 117 of the Succession Act 1965, nor with the treatment of adult children in the rather different family protection jurisdictions of England and Northern Ireland. An analysis of the exercise of judicial discretion in these jurisdictions would form the basis for separate papers. Yet the experience of the Commonwealth courts is clearly not without relevance to the operation of the flexible restraint on testamentary freedom closer to home. While the inheritance law of any given society will be shaped and influenced by its distinctive cultural requirements, certain problems are common to all succession regimes. The place which adult children should occupy within the framework surely falls within this latter category.

NINETY YEARS OF MORAL DUTY OF A WISE AND JUST FATHER

Introduction

Restraints of space preclude a systematic survey of the case law from each of the Commonwealth jurisdictions which have granted *locus standi* to adult children and measured the testator's conduct in such cases against the standard of a "wise and just father". The approach adopted below, therefore, has been to focus primarily on the New Zealand jurisprudence, in that the principles first developed there have then often been endorsed and applied by other Commonwealth jurisdictions. In order to illustrate both the huge range of factors which the courts have come to consider as relevant to the testator's moral duty and the major trends in the decades since Edward J's judgment in *Allardice* was handed down, several Canadian and Australian cases have also been included. Little attempt has been made, however, to deal with the jurisdictions separately – it is generally the case that Commonwealth countries "have drawn freely on one another's decisions"⁵⁶ and differences of interpretation between minor variations of wording have been discouraged:

"The legislation of the various [Australian] States is grounded on the same policy and found its source in New Zealand. Refined distinctions between the Acts are to be avoided."⁵⁷

The First Half Century

Many cases have expressed the sentiment that what constitutes a "breach of moral duty" will change from period to period because a family protection statute "is a living piece of legislation", the application of which "must be governed by the climate of the time."⁵⁸ Change during the first half of the century, however, was relatively slow. Rather, this period was marked by a number of gradual developments, minor in themselves, but which built on

⁵⁶ Laufer, *op cit* n 9 at 288. Laufer also notes that this characteristic has not been shared by the English judiciary. However, the English (and Northern Ireland) legislation is less broadly drawn. The implications of the issues discussed in this paper for England, Wales and Northern Ireland are discussed briefly below.

⁵⁷ *Per* Dixon CJ in *Coates v National Trustees Executors and Agency Co Ltd* 95 CLR 494 at 507.

⁵⁸ *Per* McCarthy P in *Re Wilson* [1973] 2 NZLR 359 at 362.

the foundations laid in *Allardice* and paved the way for the broader interpretation of the statutory provisions which would not properly gain momentum until the latter half of the twentieth century.

As noted earlier, the first decisions of the New Zealand Courts on the Testator's Family Maintenance Act were characterised by a very narrow interpretation of the words "adequate provision" for "proper maintenance", one which cast the function of this legislation firmly in terms of support for those in need, and then only for those whom the deceased had a lifetime obligation to support. Thus the link with the Destitute Persons Act which had figured prominently in the Parliamentary debates continued to provide the predominant thrust of the pre-*Allardice* judgments, most of which (including those delivered by Edwards J and Stout CJ) limited the Testator's Family Maintenance Act in this manner. On occasions the amount to which an applicant would have been entitled under the Destitute Persons legislation was also used to determine the *quantum* of the award to be made under the Testator's Family Maintenance Act.⁵⁹

It was emphasised earlier that *Allardice's* full impact cannot be assessed in the short-term. It would be doing the decision an injustice, however, not to recognise that it made an immediate contribution to the interpretation of its own "moral duty" test, beyond the basic principle that adult daughters married to men of poor means were owed obligations which were more extensive than those owed to their unmarried brothers. It also provided clear appellate authority that family protection was not simply an extension of the Destitute Persons Act. Hence, within just ten years of the enactment of the Testator's Family Maintenance Act, an Act which had secured its passage through the New Zealand Parliament by virtue of the savings it was promising the welfare purse, there had been unequivocal rejection of the idea that family protection was intended to serve only a support function in its narrowest sense.

Notwithstanding this, however, or the dicta in *Allardice* that the size of a deceased's estate and certain other matters may be relevant to the assessment of any moral duty, the emphasis remained for some years on the *economic need* of the applicant, albeit no longer need in the narrow context of the Destitute Persons Act. It was twelve years before there was any real evidence of a more liberal approach in which courts were prepared to attach significant weight to the means of the testator. Then in *Allen v Manchester*,⁶⁰ in what has been described as a "classic formulation", Salmond J divided family protection cases into two broad categories. First, those where:

"owing to the smallness of the estate and to the nature of the testamentary dispositions, the applicant is competing with other persons who also have a moral claim upon the testator⁶¹. . . [and secondly, those] in which owing to the largeness of the estate or the nature of the testamentary dispositions, the applicant for relief is complaining not of the unjust distribution of an inadequate fund among dependants all of whom had a

⁵⁹ See *eg, Re Rush* (1901) 20 NZLR 249 at 254.

⁶⁰ [1922] 41 NZLR 218.

⁶¹ *Ibid* at 221

moral claim upon the testator, but of the failure of the testator to make out of the abundance of his resources a provision sufficient for the proper maintenance of the claimant.”⁶²

This latter type concerned:

“the more difficult function of determining *the absolute scope and limit of the moral duty* of a wealthy husband or father to make testamentary provision for the maintenance of his widow and children.”⁶³

Allardice, opined Salmond J, fell within the first category and as such was not a decision that a testator owed no moral duty to an adult son capable of earning his own living, nor that the moral duty towards an adult daughter was limited to provision of an annuity. Rather, that particular case was:

“merely a decision as to the just and proper method of distributing between the various claimants an income not adequate to meet in full the moral claims of all of them”

In *Allen v Manchester* itself the court was not required to make a decision about adult children, but Salmond J’s *dicta*, particularly his reference to the “absolute scope and limit” of the moral duty of a wealthy father towards his children, undoubtedly pointed towards significantly more substantial provision for adult offspring than that previously recognised by the courts. Within two years, in *Welsh v Mulcock*,⁶⁴ Salmond J enjoyed a further opportunity to provide more guidance on the extent of this duty. Having intervened during the submissions to ask counsel whether the word “adequate” should be considered “ethically or economically”, the learned judge made it clear during his judgment that he preferred the former approach. On a narrower level, the case is authority for the principle that the true measure of a testator’s moral obligation is not necessarily found in the standard of maintenance and way of life to which the family was accustomed in his lifetime, so that he

“cannot, by failing in his duty to [his wife and children] while alive, reduce the standard of his testamentary obligations towards them on death.”⁶⁵

However, it was *Bosch v Perpetual Trustees*⁶⁶ (an appeal from New South Wales in which, it will be recalled, the Privy Council expressly affirmed the *Allardice* moral duty approach), which heralded the most significant broadening of approach. In a clear departure from the earlier case law, which tended to focus on the meaning of the term “adequate provision”, the Board concentrated primarily on the interpretation of the words “proper maintenance”. While a small sum might be sufficient for the “adequate” maintenance of a child, Lord Romer pointed out,⁶⁷ it might not, having

⁶² *Ibid* at 222

⁶³ *Ibid* at 222. Author’s emphasis.

⁶⁴ [1924] NZLR 673.

⁶⁵ *Ibid* At 686.

⁶⁶ [1938] AC 463.

⁶⁷ *Ibid* at 476. In *Bosch* the Privy Council allowed the costs of an Oxbridge education to come within the concept of “maintenance, education and

regard to that child's station in life or the size of the estate, be sufficient for his "proper" maintenance. In particular, his Lordship continued, where the estate was large, the court would be justified in making provision to meet contingencies which might have been disregarded had the estate been small.

Bosch thus established conclusively that an applicant's economic need was only one factor in what constituted "proper" maintenance for him or her, and thus only one factor relevant to the assessment of the deceased's moral duty to that person. Not only therefore was the wording of the various family protection jurisdictions sufficiently general to accommodate the tensions between the support and contribution functions of family protection, it could be construed as reflecting those very tensions in its drafting. As Gresson J would later observe in *Re Strawbridge, Telfer v Strawbridge*:⁶⁸

"[i]t would seem that 'adequate' has regard rather to the necessities of the applicant but that what is 'proper' has regard to all the attendant circumstances."

In practice, however, the courts generally remained reluctant to make full use of the approach developed in *Bosch*⁶⁹ and the wise and just father of the middle years of the century could generally prepare to leave this world content that only children unable to support themselves because of their age, poor health or other limited circumstances could successfully challenge his own distribution of his property. By now there had developed a body of case law based on relatively well-defined judicial principles, so that notwithstanding the element of uncertainty which is inherent in any discretionary system, it was generally possible to predict with considerable confidence how cases would be decided. Writing in 1941 the editor of the *Australian Law Journal* was able to commend

"[the] Dominion Equity Judges [for giving] the jurisdiction a settled body of practice which made it possible to advise clients with accuracy as to claims under the legislation."⁷⁰

Similarly, Joseph Laufer, having reviewed authorities from Australasia and the five Canadian Provinces which had by 1955 adopted family protection mechanisms, concluded that:

"[t]he Courts have responded to [the] broad delegation of control over private property rights with characteristic self-restraint. They have emphasised the limitations on their functions rather than stressed their powers."⁷¹

The flexible restraint on testamentary freedom had been widely acknowledged as a success and many jurisdictions were starting the process of replacing, or amending, the original legislation by provisions which generally:

advancement" found in s 3 of the Testator's Family Maintenance and Guardianship of Infants Act 1916.

⁶⁸ [1952] GLR 442 at 445.

⁶⁹ See further Laufer, *op cit* n 9; Peart, *op cit* n 28 and M Brown, "The Dependents Relief Act" (1940) 18 *Canadian Bar Review* 261 at 264.

⁷⁰ (1941) 15 *Australian Law Journal* at p 198.

⁷¹ Laufer, *op cit* n 9 at p 289

“tended. . . to discard limitations, both substantive and procedural, by which anxious legislatures had originally sought to contain both the scope of the statutes and, especially, the sweep of judicial discretion.”⁷²

For instance, in New Zealand the Family Protection Act 1908 was repealed and replaced by the Family Protection Act 1955⁷³ which widened the class of eligible applicant. However, the key provision remained substantially unchanged; had the testator provided “adequate” provision for the “proper” maintenance of the applicant. Breaches of moral duty by the just and wise father, although not incorporated in any of these statutes, remained the standard against which the court would determine whether or not this had been achieved. While, however, these statutory changes were not directly related to the scope of discretion, their expansive nature may have acted as a catalyst for a similarly expansive approach to the interpretation of the legislation, for it was around this time that there was a perceptible change - a change which was especially evident in relation to adult child claimants.

Mid-Century to Date

The certainty, predictability and consistency which were widely accepted concerning the first generation or so of family protection cases, even at the “sharp” end of its jurisdiction, would appear to be no longer a feature of the modern Commonwealth jurisprudence in the area. Thus while a future member of the New South Wales Bench could confidently proclaim in 1951 that:

“no judge has ever worn the mantle [of just and wise fathers] with the abandon one would expect of a Medieval Chancellor suitably modernised. . .”⁷⁴

one wonders if the proverbial “Chancellor’s foot” is not again very much in evidence:

“Emphasising the testator’s ‘moral duty’ leads to a judicial free-for-all. Alternatively one might say, in a free adaptation of the words of John Selden, that with such judicial power a testator’s ‘moral duty’ will vary according to the conscience of each individual judge, and as that is longer or narrower, so is the duty.”⁷⁵

Moreover, this appears to have been the experience, to a greater or lesser extent, of every jurisdiction in which the “moral duty” approach has been

⁷² *Ibid* at p 284.

⁷³ The legislation from the other jurisdictions was noted above at n2.

⁷⁴ C McLelland, *op cit* n 52 at 344.

⁷⁵ Oosterhoff, “Succession Law in the Antipodes: Proposals for Reform in New Zealand” (1997) 16 *E & TJ* 230 at 236. The famous words of John Selden referred to are: “Equity is a roguish thing. For law we have a measure . . . Equity is according the the conscience of him that is Chancellor, and as that is longer or narrowe, so is equity. Tis all as if they should make the standard for the measure a Chancellor’s foot.”

embraced and, as might be expected, the difficulties have been particularly acute in relation to able-bodied adult children.⁷⁶

In British Columbia, one judge has observed in relation to such claims under the Wills Variation Act 1979 and its predecessor:

“The law developed in these cases is in a state of disarray such that it is all but impossible to predict with any confidence what result will flow from any given set of facts and unduly difficult to decide what the result should be. . . .”⁷⁷

and a leading academic has despaired that the courts of that Province appear to be “dispensing palm tree justice,”⁷⁸ likening the Wills Variation Act to “a ship adrift at sea without an accurate compass or dependable rudder.”⁷⁹ Similarly, in Manitoba, the Law Reform Commission,⁸⁰ having cited numerous examples of successful claims by able-bodied adult children⁸¹ who had not been dependent on the testator for support, criticised the “uncertainty and inconvenience” of the moral duty approach to the Testator’s Family Maintenance Act, a further effect of which was to transform family protection

“in substance. . . from a mere limitation on testamentary power into an emerging principle that children are entitled to a share of their parent’s capital estate.”⁸²

Even in New Zealand, the country which gave the world both the discretionary-model family protection mechanism and the juristic father who has been largely responsible for shaping its course, widespread dissatisfaction, particularly with the state of the authorities on adult children, has led the New Zealand Law Commission to undertake a fullscale review of all existing succession structures.⁸³ A survey commissioned as part of this review,⁸⁴ by New Zealand academic Nicola Peart involving 235 cases from 1984-1995 where children claimed against their parent’s estates under the Family Protection Act 1955, suggested that as many as 91.5% of children’s

⁷⁶ There have been other controversial categories of eligible applicant in a number of jurisdictions - *eg* those who claim by virtue of *de facto* dependency and cohabitants. However, it is probably a fair comment that they have not generated the same difficulties as adult children. In any event they raise somewhat different issues, and are outside the scope of this paper.

⁷⁷ *Per* Esson JA in *Price v Lypchuk Estate* 26 ETR 259 at 276.

⁷⁸ G Bale, “Palm Tree Justice and Testator’s Family Maintenance – The Confusing Saga and Uncertainty in BC Courts” 26 ETR 295 at 313

⁷⁹ *Ibid* at 315.

⁸⁰ Manitoba Law Reform Commission Reports, *Report on Testator’s Family Maintenance Act* (1985). Following the recommendation of this report the Testator’s Family Maintenance Act 1946 was replaced by the Dependents Relief Act 1989-90 under which applications must be grounded in economic need.

⁸¹ *Eg Re Bartel Estate* (1982) 16 Man R (2d) 29 (QB); *Dutka v Dutka’s Estate* (1980) 7 Man R (2d) 211 (QB); *Menrad v Blowers* (1982) 16 Man R (2d) 288.

⁸² *Op cit* n 80 at p 26.

⁸³ In August 1996 the Commission published a detailed technical paper - *Succession Law: Testamentary Claims - A Discussion Paper* (NZLC PP 24, 1996). This was followed by a Report in August 1997 which included a draft Succession (Adjustment) Act. It has not yet been enacted.

⁸⁴ Peart (1995) *BFLJ* 224.

claims are successful when contested. In more than a quarter of these the children had not been dependent on their deceased *parent* (the just and wise *father* having been suitably modernised by the 1990's). As Peart has concluded:

“[a] substantial number of adult children apply under the Family Protection Act every year alleging a breach of moral duty on the part of their deceased parent and they are almost always successful . . . The testatmentary freedom which exists now is but a hollow image of its former self, [and]. . . exists in reality only for those without families.”⁸⁵

In the course of the last forty years the Commonwealth courts have generally overcome their initial reluctance to wear the mantles of just and wise fathers with “abandon” (the suggestion by one Canadian commentator, that this early reluctance was rooted in the traditional common law separation of law and morals, would seem to have much merit)⁸⁶ to move to a position where, although the economic need of the applicant remained the first consideration, other ethical issues were then introduced and taken into account, through to one in which an applicant may succeed even if he or she is in very comfortable financial circumstances:

“The ‘need’ of an applicant, or rather his or her needs – the plural form is I think preferable – cannot be considered *in vacuo*. What has to be assessed are the merits of the claim having regard to the applicant’s circumstances as at the date of the death of the testator; relations between the testator and the applicant in the past; and the extent of his estate and the strength of other claims”.⁸⁷

The above quote came from *Re Harrison*, a case in which it was held that a testator could owe a moral obligation to an adult daughter even though she was not in necessitous circumstances; indeed, even an adult in a reasonably strong financial position could expect to receive something under the will in view of the size of the testator’s estate. In *Harrison* itself the court was clearly influenced by the source of the deceased’s money, much of which he had inherited from his own father and an uncle – it was “grandfather’s money”. The testator’s widow was not the mother of the applicant, and it was thought to be unlikely that she would make any provision for the applicant in her will. In fact, it appears that applicants who are the children of a first marriage when the testator has remarried now invariably receive something under the family protection legislation (even against the widow) in both New Zealand and Australia⁸⁸ – a clear example of the statutes serving elements of both the “fair distribution”, and “dynastic” functions, but certainly not a support function.

Another situation in which adult children appear to succeed, virtually as of right, to a share of the estate, at least in Australia, is where the testator was

⁸⁵ Peart, *op cit* at n 28 at 209.

⁸⁶ M Brown, *op cit* at n 69 above at 264.

⁸⁷ *Re Harrison* [1962] NZLR 6 *per* Gresson J.

⁸⁸ See generally Peart, *op cit* n 28 and Atherton and Vines, *Australian Succession Law, Commentary and Materials* (Australia, 1996) 689-705.

very wealthy.⁸⁹ Even in relation to more modest estates, however, the likelihood of securing an award is much better than it would have been in past decades. The principle of self-reliance which pervaded the earlier jurisprudence concerning able-bodied applicants, particularly males, has generally given way to a more sympathetic approach. If this change in judicial attitude can be marked by a single turning-point, it is probably the Australian decision of *Hughes v National Trustees Executors and Agency Co of Australasia Ltd*,⁹⁰ in which Gibbs J endorsed intervention in favour of an adult son who, notwithstanding his physical well-being, simply found it more difficult than most to deal with the vicissitudes of life's rich tapestry:

"I do not think that the Legislature intended that provision under the Act should be given to those that are efficient and successful rather than to those who are not. A just father's moral duty is to assist the lame ducks among his offspring, provided that they be not morally or otherwise undeserving."⁹¹

Indeed, often the modern basis of successful child applications is no more than the fact that the parent's estate has a symbolic value. As Peart has observed of New Zealand claims:

"[the estate] has a symbolic value. . . in that it is the remaining tangible link to the parent. Complete or partial exclusion from the estate may then be seen as exclusion from the family bond. The estate becomes the symbol of the parent-child relationship, and it is this, rather than financial need, which appears to motivate many adult children to claim against the estate of their deceased parent."⁹²

In support of her submission that New Zealand is moving "towards a concept of family property,"⁹³ Peart charts a shift from "support stage" through "contribution stage" to "relationship stage", citing as evidence of the final stage a number of adult applicants who appear to have benefited under the Family Protection Act simply by virtue of the genealogical link with their parents:

"[children's] needs are not in the strict sense for maintenance or monetary provision but are rather to recognise the family connection and the relationship, the familial demand."⁹⁴

This more expansive exercise of discretion has been allied to a correspondingly more generous interpretation of the "disentitling provisions"; for example, in *Re Mercer*⁹⁵ Roper J noted that conduct which may have been sufficient to disentitle an applicant in earlier times would not necessarily do so in today's climate and duly made an award to a rebellious

⁸⁹ Atherton and Vines, *ibid* at 702.

⁹⁰ (1979) 143 CLR 134.

⁹¹ *Ibid* at 139.

⁹² N Peart, *op cit* n 28 at 207.

⁹³ Peart, "Towards a Concept of Family Property in New Zealand", *International Journal of Law, Policy and the Family* 10 (1996) 105.

⁹⁴ *Per* Greig J in *Re Adams*, HC Wanganui, A34/84, 1 April 1986 cited in Peart, *op cit* at n 28 at 207.

⁹⁵ [1977] 1 NZLR 469.

twenty-four year old son of the testator, on the basis that his father “should have recognised that his son’s make-up was such that he would need some support.” It would seem, therefore, that the judiciary are slow to judge children “morally or otherwise undeserving”, and that the judicious parent is now “not only wise and just but also contrite and forgiving.”⁹⁶

In short, the term “moral duty”, first clearly enunciated in *Allardice*, has come to encompass a host of different factors, including: the age, marital status, health and earning capacity of the applicant; his or her past standard of living - and future prospects; the poor conduct of the testator towards the applicant; past services provided to the deceased by the applicant ; and even simply the existence of a family link between the applicant and the deceased. The result is that both this duty, and the consequent invasion which it has made on testamentary freedom, are far more extensive than anything envisaged by either the framers of family protection legislation or the members of the New Zealand Court of Appeal who delivered the *Allardice* judgment. Successful claims by able-bodied adult children who were not in need, even in the broadest sense of the word, appear to abound in every Commonwealth jurisdiction in which the courts have perceived their function in relation to family protection as rectifying the moral failings of a wise and just father. Illustrations can readily be found of awards made to adult children which serve a support function by making provision for those in clear economic need; of others which demonstrate a restitutive or equitable distribution function, by rewarding the child’s good conduct or compensating for the testator’s bad conduct; and of yet others which further, if not a dynastic, at least a “family property” function, by symbolising the family bonds which ought to exist in a household. It is the household, it seems, of “the paragon of wisdom and justice who is seldom encountered on the Clapham omnibus”,⁹⁷ a household which now includes the ne’er do well, the lame duck and even the black sheep.

INTO A NEW CENTURY – WHAT FUTURE FOR THE WISE AND JUST FATHER?

Allardice v Allardice was the first case to analyse New Zealand’s still young family protection jurisdiction in terms of the moral duties owed by a just testator to his wife and children. Since then, notwithstanding occasional attempts by certain members of the judiciary to remind their colleagues that although

“[the notion of the moral duty owed by a wise and just testator] may be a helpful gloss in many cases, . . . when a critical question of meaning arises, the question must be answered by reference to the text, and not by reference to the gloss. It is the text, and not the gloss, that we are called upon to interpret. . .”⁹⁸

⁹⁶ J L Caldwell, “Family Protection Act 1955 - Moral Duty and Adult Children” (1982) *New Zealand Law Journal* 215 at 218.

⁹⁷ *Re Bartel Estate* (1982) 16 Man R 29 at 32, per Scollin J.

⁹⁸ Per Fullagar J dissenting in *Coates v National Trustees Executors and Agency* (1956) 95 CLR 494 at 523. See also Jordan CJ in *Re FJ McNamara dec’d* (1938) WN (NSW) 180 – the gloss “likely to obscure rather than to clarify the

Edwards J's words which constitute this "gloss" have become the *locus classicus* in the area, repeated

"every year since 1910 . . . in every Court in [New Zealand], and never with any qualification or question. The principle which they expound has again and again been approved by our greatest Judges. . ."⁹⁹

– a remark which is equally valid for the Australasian and a number of the Canadian Courts. Few would dispute that Cooke J was correct in his assessment that this judge-made creation has become "too deeply embedded"¹⁰⁰ in family protection jurisprudence in these countries to be removed other than by statutory intervention.

Yet it would appear that the experience of those Australasian and Canadian courts which have perceived the primary purpose of family protection statutes as enforcing the moral duties of the deceased has not been an unmitigated success, at least in relation to adult child claimants. On the contrary, the evidence from these jurisdictions suggests that the law in this area has been racked with uncertainty and inconsistency, to such a degree that it has become virtually impossible for anyone with a legitimate interest to know the exact nature of either their obligations or their rights. Meanwhile, this level of unpredictability provides a very fertile breeding ground for that inevitable contra-indication of any discretionary system - the frivolous claim. It might be impossible to eliminate this type of action completely, but amid such confusion even the most unworthy of applicants are endowed with that "enhanced nuisance value"¹⁰¹ which tends to make any settlement more attractive to a beneficiary under a will than the risk of litigation.

It would, of course, be doing an injustice to lay responsibility for this state of affairs entirely on the shoulders of the judiciary. No doubt when Edwards J expounded his now familiar words it was a *bona fide* attempt to fill the very conspicuous void left by the legislature. As has been noted, there was a certain ambiguity about the objectives of the seminal legislation, New Zealand's Testator's Family Maintenance Act 1900, while its key substantive provision (that subsequently adopted by legislatures throughout the Commonwealth) had a general elasticity which could accommodate different and potentially conflicting interpretations - a tension which was heightened by the inclusion of the terms "adequate" and "proper" in virtual juxtaposition. While, however, the absence of clear direction from the legislature has done nothing to ameliorate the burden on successive generations of judges as they have embraced the task of defining the scope of the family protection jurisdiction, Gordon Bale's observation that "proper maintenance is a much less amorphous concept than a fair share of the

legislation"; *Hughes v National Trustees Executors and Agency* (1979) 143 CLR 134; *Goodman v Windeyer* (1980) 144 CLR 490; *Singer v Berghouse* (No2) (1994) 123 ALR 481.

⁹⁹ *Per* Turner J in *Re McGregor* [1961] NZLR 1077.

¹⁰⁰ In *Re Z* [1979] 2 NZLR 495 at 506.

¹⁰¹ Law Reform Commission of British Columbia – *Report of Statutory Succession Rights* (1983) Report No 70– reservation by Mr Arthur Close at p 153.

estate”¹⁰² is still well made. Whatever the attraction of any judicial gloss, it must never be allowed to detract from the requirement that the initial question to be answered in every application is a *jurisdictional* one (that is, did the deceased make adequate provision for the proper maintenance of the applicant?). It seems, however, that as the judiciary have placed more and more emphasis on assessing the “moral duty” of the testator, they have shifted their focus completely from the actual statutory provisions which they are called to interpret – after all

“[f]lights of judicial rhetoric [such as the moral duty of a wise and just husband and father] lend themselves to quotation more than colourless prose - so much so that the statutory language is forgotten.”¹⁰³

The unfortunate result is that the function of the legislation, admittedly never as transparent as it might have been, has become totally obscured.

It is submitted that this state of affairs cannot be remedied simply by enshrining the “moral duty” test in statute, as in the Republic of Ireland. Breach of moral duty is simply not the most satisfactory basis for judicial intervention in discretionary family protection systems, and is set to become even less satisfactory as we progress into the twenty-first century.

In the first instance, the concept itself is inherently too subjective, and will inevitably remain so – no matter how often judges declare, as, for example, the Irish judiciary have,¹⁰⁴ that they are applying the test “objectively”. Any given individual’s (and every judge’s) conception of moral duty will vary depending on factors such as age, social background, personal experience and religious beliefs. Furthermore, the Commonwealth courts have endorsed an “ambulatory” approach to the concept, acknowledging that changing social attitudes will have an influence on the existence and extent of moral duties.¹⁰⁵ For example, in Australia it has been said that the judiciary are required:

“to apply prevailing community standards of what is right and appropriate since it is by those standards that the content both of the moral duty owed by a just husband and father to his wife and children and of departures from it will be measured.”

One wonders, however, whether such a “community standards” approach is realistic for the present multi-cultural postmodern societies, societies which are characterised by ever more complex social relationships, successive marriages, *de facto* unions and reconstituted families. It was surely no coincidence that it was in the 1960’s that the Australasian and Canadian Courts first experienced major difficulties in applying the moral duty test consistently.

¹⁰² G Bale, *op cit* at n 78 at p 314.

¹⁰³ G Bale, *op cit* at n 78 at p 316.

¹⁰⁴ See *eg* Keane J in *Re JR* (November 1979, unreported).

¹⁰⁵ *Re Wilson op cit* at n 58. See also *Little v Angus* [1981] 1 NZLR 126 at 127 (NZ CofA). In relation to Canada, see Esson JA in *Price v Lypchuk Estate* 26 ETR 259 at 268 – one must discharge one’s parental and marital duties “in accordance with a contemporary view of testamentary independence”.

Secondly, one must consider the much wider debate about whether it is appropriate for the law to be enforcing familial moral obligations, rather than leaving such complex issues where they were historically, “within the purview of theologians, philosophers and the individual conscience”.¹⁰⁶ Back in 1940 Canadian Mannie Brown analysed the judiciary’s adherence to the *Allardice* moral duty approach as part of what he considered to be a much broader trend during the first half of this century of converting moral obligations into legal obligations whenever possible.¹⁰⁷ At the cusp of the new century, however, a similar approach seems neither desirable nor desired – as Grainer has argued in an article based on her submissions to the New Zealand Law Commission,¹⁰⁸ moral duties are often most effectively controlled by social pressure.

Not unrelated to the earlier point that moral duty is inherently subjective, a recent study carried out by Finch and Mason in England¹⁰⁹ suggests that it is virtually impossible to build social norms about family responsibility into inheritance laws, for the simple reason that there is no broad consensus regarding such responsibilities. Using a series of vignettes, Finch and Mason focused in particular “on the question of whether people generally endorse the view that the right to inherit should be based simply on genealogical ties, unconnected with any other patterns of responsibility within families (a model which we called the ‘black box’) or whether people generally think that the right to inherit to an extent can be earned through special commitments acknowledged and fulfilled.”¹¹⁰ Although some of their findings supported the ‘black box’ approach, that is, inheritance is not something to be earned, there were also indications that the idea of an implicit bargain within the family had not been completely rejected.

Finally, and most significantly, however, the fundamental problem in relation to adult children has been the failure of legislatures to define precisely on what basis the judiciary are to interfere with the deceased’s testamentary provision and make an award in an applicant’s favour. This problem cannot be addressed merely by enshrining the moral duty standard in statute, for breach of some undefined moral duty can never, by itself, provide a sufficiently rational foundation for a family protection jurisdiction. While purporting to rectify alleged failures of a deceased’s parental duties the Australasian and Canadian Courts have been furthering one or more of several very distinct functions. Examples (a non-exhaustive list) include: providing barely subsistence level support; compensating for the deceased’s lifetime of neglect; rewarding past services; remedying the deceased’s procrastination in relation to updating his will; reimbursing the applicant’s

¹⁰⁶ Manitoba Law Reform Commission Report, *op cit* at n 80, at p 23.

¹⁰⁷ M Brown, *op cit* at n 69 at 264.

¹⁰⁸ V Grainer, “Is Family Protection a Question of Moral Duty?” (1994) 24 *VUWLR* 141.

¹⁰⁹ See generally J Finch and J Mason, *Negotiating Family Responsibilities* (1993) and J Finch and J Mason, “Family Responsibilities and Inheritance in Great Britain”, in M Maclean and J Kurczewski (eds), *Families, Politics and the Law: Perspectives for East and West Europe* (1994) Chapter 6, pp 97 -121. Grainer, *ibid* also referred to this research in her New Zealand context.

¹¹⁰ *Ibid*, at 116.

expenditure; acknowledging a blood relationship or a genealogical tie. While the courts in the Republic of Ireland have generally been unprepared to interfere with a testator's wishes quite so readily¹¹¹ as their Commonwealth counterparts, illustrations can be found within the jurisprudence on section 117 of the Succession Act 1965 of awards made which further several of the objectives on the above list, for example: the reward of past services,¹¹² compensating for past neglect¹¹³ and acknowledging the family relationship.¹¹⁴

During the 1970's and 1980's various Commonwealth law reform agencies undertook reviews of the operation of family protection legislation within their jurisdictions, including, *inter alia*, the position of mature child claimants. This process focused welcome attention on the very diffuse purposes which had been underpinning children's claims on their parents' estates. The response of both the Alberta Institute of Law Research and Reform¹¹⁵ and the Law Reform Commission of Manitoba¹¹⁶ (recommendations which were subsequently enacted)¹¹⁷ was that family protection should be restricted to a support function and, on the basis that duties enforceable *post mortem* should not be inconsistent with those duties legally enforceable during the testator's lifetime, eligibility should be limited to minor children or those under twenty-three still in education or vocational training. However, elsewhere the broader moral duty approach was generally endorsed, expressly or implicitly, and there was little attempt to clarify the purposes or objectives of the legislation.¹¹⁸

Unless, however, the law defines clearly and precisely why a family protection award is being made in favour of a particular applicant and what function it is serving, the very future of the flexible restraint on testamentary freedom, a mechanism which, properly guided and controlled, can make a very useful and effective contribution to modern succession law, must be in

¹¹¹ See generally, Brady, *Succession Law in Ireland*, (2nd ed, 1996) and for a more analytical approach to the earlier cases see Cooney, "Succession and Judicial Discretion in Ireland: the Section 117 Cases" (1980) *Irish Jurist* (ns) 66. A striking feature of the jurisprudence on s 117 is the dearth of references to Commonwealth authorities – a phenomenon for which Kenny J is no doubt partly responsible with his comments in *Re G M* [(1972) 106 *ILTR* 82] that such cases were of little assistance to the Irish judiciary because the legislation which they were interpreting had a much broader remit than its Commonwealth equivalents in that it was not limited in its application to children who were dependent upon the testator. However, in light of the authorities discussed above, one must doubt the accuracy of this appraisal.

¹¹² *In the Estate of Matilda McGarry, dec'd, McGarry v Byrne* (November, 1978, unreported).

¹¹³ *HL v Governor and Company of the Bank of Ireland*, (July 1978, unreported).

¹¹⁴ *In the matter of the Estate of J R* (November 1979, unreported).

¹¹⁵ Family Relief Report No 29, 1978.

¹¹⁶ *Op cit* at n 80.

¹¹⁷ In Alberta by the Family Relief Act 1980; in Manitoba by the Dependents Relief Act 1990.

¹¹⁸ See, *eg*, Law Reform Commission of British Columbia, *op cit* at note 100 (subject to a powerful dissent by Arthur Close) and Law Reform Commission, New South Wales, *Report on the Testator's Family Maintenance and Guardianship of Infants Act* (1977).

doubt. A concept such as “breach of moral duty” (or indeed, a “fair distribution of the estate”) is too vague and nebulous to ensure that the purposes and effect of the law are clearly communicated to all concerned and meanwhile members of the judiciary are at liberty to hide behind this rhetoric, giving frequent assurances that they have no power to recast unjust wills to effect a more equitable distribution - and then proceeding to do just that. In light of the current state of the authorities in some of the jurisdictions considered above, it is no exaggeration that *Allardice*, as subsequently applied, has transmuted the right of an adult child to ask for “proper maintenance” from his or her deceased parent’s estate, into a right to receive a fair share of that estate in the absence of any direct evidence of justification for disinheritance. Such judicial metamorphosis of a statutory provision without the sanction of the legislature clearly goes beyond the proper role of the court in relation to either the interpretation and application of statutes, or the exercise of discretion.

NEW ZEALAND – LEADING THE WORLD ONCE MORE?

Any clarification and subsequent reformulation of the objectives of family protection legislation cannot, however, be undertaken in a vacuum, but must take place in tandem with an appraisal of other legal mechanisms through which persons may secure a share of a deceased’s estate. For instance, if family protection for child applicants is to be restricted to a support function there undoubtedly will exist a “gap” – even jurisdictions which have adopted this narrow approach, such as Manitoba, recognise that non-dependent adult children may sometimes be entitled to part of an estate, for example, where a child has worked for his or her parent for virtually no wages. How should these issues be addressed – through the general law of trust, contract and estoppel, or by the introduction of a specific statutory remedy?

With its draft Succession (Adjustment) Act,¹¹⁹ the most serious attempt to date to overhaul and streamline the entire area of testamentary claims, New Zealand again appears to be leading the common law world in family inheritance matters. This Act, currently being considered by the New Zealand Parliament, proposes the abolition of the moral duty test and the introduction of two distinct types of claim - support claims, recognising *dependency* on the testator which should continue beyond the grave, and property division/contribution claims. Children would be able to make a support claim (to be supported for their reasonable needs in life) if they are minors, or under twenty-five and undertaking educational or vocational training, or unable to earn a reasonable, independent livelihood because of a physical, intellectual or mental disability which occurred before they reached twenty-five.¹²⁰ Following considerable consultation and discussion, it has been proposed that adult independent children may also make a support claim, in their case described as a “needs claim”, to provide them with the necessities of life – but only where they are genuinely in need and it is

¹¹⁹ In August 1996 the New Zealand Law Commission published *Succession Law: A Discussion Paper* (NZLC PP 24, 1996). This was followed by a Report, *Testamentary Claims* (Report 39) published in August 1997 and containing a draft Succession (Adjustment) Act. This Act has not yet been enacted.

¹²⁰ Draft Succession (Adjustment) Act, s 27.

possible, without unfairness to those otherwise entitled to the estate, to provide periodic payments sufficient to alleviate their need.¹²¹

In addition, it is proposed that children (and anyone else) may make a contribution claim where they are seeking the return of unremunerated benefits which they have conferred on the testator. Such claims, which are designed to replace the existing “mosaic of statute law, common law and equity”¹²² (namely contract, trusts, estoppel, restitution and the Law Reform (Testamentary Promises) Act 1949, New Zealand’s unique statutory machinery for determining claims by persons who have contributed valuable benefits to the deceased), will be able to be brought where the applicant provided a benefit to the deceased during his or her lifetime and either the deceased expressly promised to make provision for the applicant in return for the benefit, or it is unjust for the estate to retain the benefit without making provision for the applicant.¹²³ It is impossible to discuss these proposals in detail in this article, but suffice it to say, at this juncture, that any attempt to separate out the various functions of testamentary claims can only be a significant improvement on the current situation.

Before concluding it may be appropriate to mention briefly two family protection jurisdictions which have thus far been conspicuous by their absence, namely England and Northern Ireland. Their exclusion is justified by the subject-matter - *Allardice’s* judicious father has not been invoked, at least overtly, by either the English judiciary when determining whether a testator has “failed to make reasonable financial provision” for the “maintenance” of his children under section 3 of the Inheritance (Provision for Family and Dependents) Act 1975, or by their Northern Ireland counterparts when exercising their comparable jurisdiction which is conferred by article 5 of the Inheritance (Provision for Family and Dependents)(Northern Ireland) Order 1979. It is interesting, however, that the existence of a “moral obligation” owed by the deceased to the applicant has been an influential factor in relation to adult child claimants in both of these jurisdictions.¹²⁴ The more objective criteria of the 1975 Act and 1979 Order have generally resulted in much less uncertainty and inconsistency in relation to adult children than that which exists in the jurisdictions discussed earlier, but evidence can still be found within the reported case law of awards made which clearly further a “fair distribution of the estate” function rather than a support function.¹²⁵ This is particularly so in Northern Ireland, where adult children have been treated more generously than in England,¹²⁶

¹²¹ *Ibid*, s 29. Such an award may only be made out of the residue of the estate. There is also a facility for adult children to seek a chattel of modest value as a keepsake from their parent’s estate (known as a “memento claim”): s 30. A quick procedure through the Disputes Tribunal is provided for such memento claims.

¹²² New Zealand Law Commission Report Number 39 (1997), *Testamentary Claims*, at p 13.

¹²³ *Ibid*, s 30.

¹²⁴ In England, see in particular *Re Coventry* [1980] Ch 461, although more recently “moral obligation” has tended to be replaced by “special circumstances” - see *Re Hancock* [1998] 2 FLR 348, and *Espinosa v Bourke* [1999] 1 FLR 747.

¹²⁵ In England see *eg Re Leach* [1986] Ch 226 and *Re Goodchild* [1996] 1 WLR 694, both of which were influenced by the source of the deceased’s money.

¹²⁶ *Re McGarrell* [1983] 8 NIJB and *Re Creaney* [1984] NI 397.

generating both confusion and unpredictability, and, to the author's personal knowledge, a number of *ex gratia* settlements made out of estates to persons whose claim under the 1979 Order could at best be described as weak. The problems which *Allardice* and its aftermath have forced New Zealand to address are not without relevance to the discretionary family protection mechanisms of England and, more especially, of Northern Ireland.

CONCLUSION

The issues which have been under discussion in this article are, of course, part of a much broader debate about the position which children should occupy in a modern family property system, in an era in which inheritance laws have been characterised by the movement away from blood relatives in favour of the spouse¹²⁷ and in which the most valuable intergenerational transfer may be an education (and the resultant earning capacity) rather than the family patrimony.¹²⁸ Indeed, the focus in recent decades on the property interests of the wife (and, of course, her *de facto* counterpart) has undoubtedly diverted interest away from non minor children. Few would dispute that a spouse's claim on the family assets, both in life and on death, is paramount (or that the law's recognition of this fact was long overdue), but there are still a number of issues relating to the interests of other 'family members' (howsoever this class should be defined in the new millennium), which also deserve to be addressed. One of these is which, if any, circumstances, duties, or responsibilities relating to an adult child should be translated into proprietary interests in his or her deceased parent's estate? And, then, by virtue of what legal mechanism - a discretionary family protection system, or otherwise? To highlight just one issue which is of increasing concern, what of carers (a group including, but obviously much wider than, adult children)? With increased longevity and the acknowledgement throughout the western world that public funds will not be sufficient to care for an elderly population - who *will* be expected to care, and what part, if any, should inheritance play in the equation?

Allardice v Allardice has undoubtedly secured its place in the history of family property. Undoubtedly, also, its legacy will continue to be felt, not least on these shores, well into the twenty-first century. The manner in which it has shaped family inheritance law is not, however, without a number of difficulties. Like New Zealand other jurisdictions might benefit from a thorough appraisal of the functions and application of their family protection legislation and, in particular, its relationship to the rest of the family property regime.

¹²⁷ See generally MA Glendon, *The Transformation of Family Law* (1989), and MA Glendon, *State, Law and Family* (1977) pp 279-280.

¹²⁸ See generally Langbein, "The Twentieth-Century Revolution in Family Wealth Transmission" (1988) 86 *Mich LR* 722.

MODERNISING THE REGULATION OF WATER POLLUTION IN NORTHERN IRELAND

PART TWO: RECASTING THE DISCHARGE CONSENT SYSTEM AND AN AGENDA FOR FURTHER REFORM

*Sharon Turner, Lecturer in Law, Queen's University of Belfast**

INTRODUCTION

Part Two of this extended analysis of the Water (Northern Ireland) Order 1999¹ will address two issues. First, it will analyse the extensive changes made to one of the core regulatory mechanisms used to prevent and control freshwater pollution in Northern Ireland: namely, the discharge consent system. Second, it will identify an agenda for further reform of the regulatory framework governing Northern Ireland's aquatic environment.

National and European Union policy on freshwater conservation has consistently emphasised the need to maintain and improve the quality of freshwater, not only because it is a vital economic resource for industry and agriculture, but also because it is a fundamental and highly valued component of the United Kingdom's environment and ecology.² Although the United Kingdom and the European Commission are both committed to developing economic instruments as alternative strategies for delivering environmental protection, administrative regulation will remain at the heart of national and European policy on the environment for the foreseeable future.³ Given that the discharge consent system plays a pivotal role in the regulation of the aquatic environment, it is vital that the legal framework governing this regulatory mechanism gives expression to contemporary

* The author would like to thank Professor Colin Reid (University of Dundee), Professor Richard Macrory (University College London), Professor Brigid Hadfield (Queen's University, Belfast) and Clive Mellon (Senior Conservation Officer, RSPB(NI)) for their comments on earlier drafts of this work. Thanks are also due to my research assistant Sharon Thompson. In addition, the author would like to acknowledge the funding provided by the Law School, Queen's University Belfast, which supported the research of this paper. Part I of this article is published in (2000) 51 *NILQ* 65.

¹ SI 1999/662 (NI 6).

² *This Common Inheritance: Britain's Environmental Strategy*, (1990) Cm 1200, Chapter 12; *Sustainable Development: The United Kingdom Strategy*, (1994) Cm 2426, Chapter 8; *Better Quality of Life: A Strategy for Sustainable Development for the United Kingdom*, (1999) Cm 4345, at pp 76-78.

³ *This Common Inheritance*, *ibid*, at p 13 (paras. 1.26-32); *Sustainable Development: The United Kingdom Strategy*, *ibid*, p 34 (paras 3.20-21); *A Better Quality of Life*, *ibid*, at pp 26-27; *Towards Sustainability*, the EC's Fifth Action Programme on the Environment, 1993-2000, OJ C138, 17 May 1993, Chapter 7; Bell, *Ball & Bell on Environmental Law*, (4th ed, 1997), Chapter 5 generally and particularly pp 117-124; Bowman, "Improving the Quality of Our Water: The Role of Regulation by the NRA" (1992) 70 *Public Administration*, pp 565-575; Weale, "Environmental Regulation and Administrative Reform in Britain", in Majone (ed) *Regulating Europe*, Routledge (1996) at pp 106-130.

policies and principles on the prevention and control of water pollution. However, in contrast to the consent system in Great Britain, which has been the subject of far-reaching reform during the past thirty years,⁴ reform of the discharge consent system operating in Northern Ireland has been considerably delayed. Not before time, the Water (NI) Order 1999 fundamentally recasts the legal framework governing the consent system in Northern Ireland and in doing so, brings this vital regulatory mechanism into line with the more sophisticated and environmentally progressive systems applying in Great Britain.

Despite the far-reaching nature of the reforms introduced by the Water (NI) Order 1999, a number of fundamental weaknesses remain at the heart of the regulatory framework governing water pollution in Northern Ireland. First, the Department of the Environment (Northern Ireland) which is responsible for administering and enforcing the system of pollution control still operates as both poacher and gamekeeper in this context. Secondly, the aquatic environment in Northern Ireland will not be afforded the degree of protection extended to freshwater in England and Wales. Thirdly, the Comptroller and Auditor General for Northern Ireland has recently raised serious questions concerning the DoE(NI)'s ability to regulate water pollution in an effective manner. To compound these problems, the Government essentially missed the opportunity presented by the making of the Water Order to lay the foundations for the implementation of the tidal wave of reform posed by the forthcoming adoption of the EC's Water Framework Directive expected later in 2000. Although each of these shortcomings is significant in its own right, collectively they will significantly undermine the legal protection afforded to the aquatic environment. In order to safeguard the future wellbeing of one of Northern Ireland's most important economic and cultural assets, further modernisation of the regulatory system is essential. Failure to do so will not only neutralise, to a considerable degree, the improvements wrought by the Water (NI) Order 1999, but also continued tolerance of comparatively lax regulation will almost certainly guarantee Northern Ireland's future as a pollution haven within the United Kingdom.

SECTION I

RECASTING THE DISCHARGE CONSENT SYSTEM

Unlike the system of water quality objectives, which was reviewed and updated in 1993,⁵ the system of discharge consents operating in this jurisdiction has not been overhauled in the twenty-seven years since its establishment under the Water Act (NI) 1972.⁶ Other than the introduction of a series of regulations during the early 1990s which required the DoE(NI) to ensure that discharge consents complied with a range of specific EC Directives,⁷ the core legal framework governing the administration and

⁴ (2000) 51 at *NILQ* 71-72.

⁵ The Water and Sewerage Services (Amendment) (NI) Order 1993, SI 1993/3165 (NI 16). For a discussion of the reform of water quality control in Northern Ireland, see S. Turner and K. Morrow, *Northern Ireland Environmental Law*, Gill and Macmillan (1997), at pp 189-218.

⁶ c.5.

⁷ S. Turner and K. Morrow, *ibid*, Chapter 5 at pp 189-238.

enforcement of consented discharges remained unchanged. Not surprisingly, the strategies and policies underpinning this important regulatory mechanism fell considerably behind its more environmentally sound counterparts applied by the Environment Agency (EA) in England and Wales and the Scottish Environment Protection Agency (SEPA). The consultation process concerning the review of the Water Act (NI) 1972 was launched in 1993; however, a further five years elapsed before the Government finally made the Water (NI) Order 1999.⁸ Despite its lengthy gestation, the new Order is not expected to come into force for at least another year, but when it does, it will fundamentally overhaul the entire legal framework governing discharge consents in Northern Ireland. Although the DoE(NI), like the EA and SEPA, will retain a wide discretion in the exercise of its new powers under the Order, the Department will now be responsible for administering and enforcing a consent system that almost completely approximates with the more sophisticated and environmentally sound controls imposed on consented discharges in Great Britain.⁹ Allied with the increased penalties for offences under the Order and DoE(NI)'s new powers to apply integrated pollution control to the most potentially polluting processes,¹⁰ Northern Ireland's new discharge consent system, if administered and enforced to its fullest extent, will afford the aquatic environment a level of protection that is compatible with the wider national and European commitment to the principle of sustainable development.¹¹

Section I of this article will examine the four key reforms made to the consent system and to assess their likely impact on the regulator, consent holders and the aquatic environment. First, the Order opens the traditionally closed relationship between consent holders, the fisheries bodies¹² and the

⁸ For a discussion of the reasons for this delay see, Part One of this article at (2000) 51 *NILQ* 65.

⁹ Transitional provisions have not yet been made for carrying over discharge consents granted under the Water Act (NI) 1972; however, it is anticipated that a consultation paper will be published on this issue once the Order comes into force.

¹⁰ Discharges from the most potentially polluting processes are subject to separate Integrated Pollution Control under the Industrial Pollution Control (NI) Order 1997 (SI 1997/2777 (NI 18)). Article 10(1) of the Water Order provides that a discharge or deposit made under and in accordance with a consent granted under the Order will not constitute an offence under article 23 of the 1997 Order. In addition, article 10(2) provides that a discharge or deposit made in accordance with an IPC authorisation granted under the 1997 Order will be a defence to a water pollution offence under article 9 of the 1999 Order. Similarly, article 10(2) provides that a discharge made in accordance with a waste management licence granted under the Waste and Contaminated Land (NI) Order 1997 (SI 1997/2778 (NI 19)) will not constitute an offence under article 9 of the Water Order.

¹¹ It should also be noted that the powers conferred on the DoE(NI) under the Water Order to control and prevent water pollution will not be reallocated to other Departments as a result of the devolution arrangements contained in the Northern Ireland Act 1998, or the Belfast Agreement.

¹² The Foyle Fisheries Commission and the Fisheries Conservancy Board for Northern Ireland. Both are non-departmental public bodies responsible for the application of fisheries legislation in Northern Ireland (The Fisheries Act (NI) 1966 and The Foyle Fisheries Act 1952). With the enactment of the Water Act (NI) 1972 and the establishment of the Environmental Protection Division within the DoE(NI) (the predecessor of the current Environment and Heritage Service),

DoE(NI) to allow for a much greater degree of public participation in decision-making concerning the administration and enforcement of discharge consents. In doing so, the Order gives expression to national and EU policy on public participation in environmental decision-making. Secondly, the Order confers on the DoE(NI) statutory powers to establish a system of discharge consent fees and thereby provides the basis for implementing the EU polluter pays principle in Northern Ireland. Thirdly, the Order amends the entire framework of control governing consented discharges to reflect Government policy on 'better regulation'. Although the overall pattern is undoubtedly one of more stringent control, the Order makes certain concessions to the need for economic stability and considerably clarifies the legal position of applicants for discharge consents, consent holders and the DoE(NI). Last, but by no means least, the Water Order gives clear expression to the principle of pollution prevention. The DoE(NI)'s new powers to serve enforcement notices (discussed in Part One of this article¹³) represent the most obvious manifestation of this principle in the context of the discharge consent system. However, the principle of pollution prevention also provides the rationale for many other changes to the Department's powers to control consented discharges – all of which will be addressed in Section I of this article.

Enhancing Public Participation in the Discharge Consent System

One of the central issues raised by the *Review of the Water Act (NI) 1972: A Consultation Paper*¹⁴ concerned the need to ensure greater opportunities for public participation in the discharge consent system.¹⁵ In this context, public participation embraces both increased rights of access to information concerning the operation of the discharge consent system and greater public notification of, and participation in, decision-making concerning discharge

the fisheries bodies came to act as agents of the DoE(NI) in the enforcement of the 1972 Act and now carry out virtually all sampling of watercourses and effluent discharges. In addition, they may also prosecute under the fisheries legislation in the event of a fish kill.

¹³ (2000) 51 *NILQ* 65.

¹⁴ December 1993. Published jointly by the DoE(NI) and the Department of Agriculture (NI). This paper set out the DoE(NI)'s proposals for amendment of the regulatory system laid down in the Water Act (NI) 1972.

¹⁵ For a discussion of the nature and scope of the concept of public participation in contemporary environmental law, see generally: *This Common Inheritance*, note 2 *supra*, at p 10 (paras. 1.14-15, 1.20 and 1.38); *Sustainable Development: The United Kingdom Strategy*, note 2 *supra*, Chapters 31 and 32; *A Better Quality of Life*, note 2 *supra*, at p 23; *Towards Sustainability*, note 3 *supra*, Chapter 3; Royal Commission on Environmental Pollution, *Setting Environmental Standards*, (1998) Cm 4053, Chapters 7 and 8; Bell, *Ball & Bell on Environmental Law*, note 3 *supra*, Chapter 7; Kramer, *Focus on European Environmental Law*, Sweet & Maxwell (1992), Chapters 5 and 14; Alder & Wilkenson, *Environmental Law and Ethics*, Macmillan (1999), Chapter 12; Somsen, (ed) *Protecting the European Environment: Enforcing EU Environmental Law*, Blackstone Press (1996); Principle 10 of The Rio Declaration on Environment and Development, adopted June 14, 1992 at Rio de Janeiro, UN Doc. A/CONF.151/5/Rev.1, (1992) 31 *ILM* 874.

consents.¹⁶ Although the Water Act (NI) 1972 established a public register, the DoE(NI) was only required to record discharge consents and conditions currently in force.¹⁷ Members of the public were entitled to consult the register at all reasonable hours but they did not have a statutory right to make copies of the information contained therein. In addition, the Act was silent as to the Department's power to exclude information from the register or to levy charges for consulting the register. The Act also provided opportunities for public participation in decision-making about discharge consents; however, once again this was limited in nature and the decision to notify the public lay entirely within the discretion of the DoE(NI). In effect, the regulation of discharge consents under the Water Act (NI) 1972 was an almost exclusive relationship between applicant or consent holder and the DoE(NI) as regulator.

In contrast, the Control of Pollution Act 1974 introduced much stronger provision for public participation in the regulation of water pollution control for Great Britain. Section 41 of the Control of Pollution Act 1974 (COPA) provided that the water pollution register must include not only applications for discharge consents and consents granted, but also samples of effluent and water, the analysis of these samples and the steps taken in consequence of the information. COPA made provision for the exclusion of information from the register on specific grounds but required the Secretary of State to include certificates authorising the exclusion in the register. Water authorities were also required to record a number of other notices served under the Act. In addition, COPA required that applications for discharge consents must be notified to the public. Although public consultation could be dispensed with in certain circumstances, the Act at least sought to control the regulators' discretion in this regard. As with much of COPA, the innovative provision for public participation was not invoked until 1985 with the result that secrecy remained, until recently, as one of the hallmarks of pollution regulation in Great Britain. By the mid-1980s mounting pressure at national and European level for public participation in environmental decision-making gradually forced the Government to embrace transparency as a key element in the United Kingdom's strategy on pollution control. Building on the innovations contained in COPA, the Water Act 1989¹⁸ introduced greater rights of public access to information concerning water pollution control and further enhanced public participation in decision-making about the aquatic environment. Further changes in this vein were introduced by the Environment Act 1995.¹⁹ By the early 1990s, public participation in environmental decision-making was accepted at national, European and international level as a key concept underlying contemporary law and policy on the environment.²⁰

Although EC Directive 90/313/EEC²¹ on freedom of access to information on the environment was implemented (late) in Northern Ireland in 1993, the

¹⁶ *Ibid.*

¹⁷ Section 9(9).

¹⁸ c. 15.

¹⁹ c.25.

²⁰ Note 15 *supra*.

²¹ OJ L158, 23 June 1990 implemented in Northern Ireland by the Environmental Information Regulations (NI) 1993, SR 1993 No. 45.

public participation provisions of the Water Act (NI) 1972 had remained unchanged since their enactment. Consequently, considerable legislative change was necessary to bring Northern Ireland into line with the much more generous public register provisions contained in the Water Resources Act 1991 (WRA)²² as amended. The need for far-reaching legislative reform was highlighted in the *Review of the Water Act (NI) 1972: A Consultation Paper*,²³ which set out a series of proposals designed to bring water pollution law in Northern Ireland into line with national provisions in this regard. It is interesting to note, however, that in some instances, the Department's proposals were more progressive than those contained in the WRA 1991. Overall, the Water (NI) Order 1999 considerably increases the public's right of access to information concerning the operation of the discharge consent system and affords greater opportunities to members of the public to influence decision-making in relation to discharge consents. Regrettably, however, the Government rejected the more progressive proposals and adopted the inconsistent requirements for public consultation contained in the WRA 1991 concerning applications for discharge consent, setting and reviewing water quality objectives and consents issued without an application. The principal proposals for amendment and the Government's response will each be addressed in turn.

Proposal No.1: Applications for discharge consents should be deemed legally invalid unless they are advertised.

The *Review of the Water Act (NI) 1972* pointed out that although the Water Act (NI) 1972²⁴ empowered the DoE(NI) to require an applicant to publish notices of the application in specified newspapers, there was little public interest in water pollution control during the 1970s except amongst organisations that were directly affected like angling and rowing clubs. Environmental protection was not only a concept in its infancy in the 1970s, in Northern Ireland, this issue was also very much eclipsed by the wider security problems that engulfed the Province during that period. Consequently, it was sixteen years before the Department started to use this power. The proposal that applications for discharge consents should be deemed legally invalid unless advertised was particularly progressive and indeed exceeded the obligation imposed on the Environment Agency. Unfortunately this suggestion received only a lukewarm response. The Water Order embraces the weakest elements of the system governing applications under the WRA 1991. Schedule 1 to the 1999 Order, like Schedule 10²⁵ to the WRA 1991, states that an application for a discharge consent must be advertised in such a manner as may be required by regulations.²⁶ However, paragraph 1(2) goes on to provide that regulations may dispense with this requirement if the Department considers it "appropriate to do so". It is not only a matter of considerable concern that the Department retains a discretion to dispense with the obligation to advertise applications for

²² c.57.

²³ Published jointly by the DoE(NI) and the Department of Agriculture (NI), December 1993.

²⁴ Section 9(3).

²⁵ Schedule 10 was amended by section 120, Schedule 22, para 183 to the Environment Act 1995.

²⁶ Para 1

discharge consents, but the Government's failure to explicitly circumscribe the Department's power in this regard is also regrettable.

In December 1996, The Control of Pollution (Applications, Appeals and Registers) Regulations 1996 came into force in England and Wales.²⁷ These Regulations set out, *inter alia*, the conditions governing the advertising of applications for discharge consents under the WRA 1991. Regulation 4 provides that the Environment Agency can dispense with the obligation to advertise an application for a consent if the application is either exempt from inclusion in the public register on the grounds of national security (discussed below), or the Agency considers that the proposed discharge is unlikely to have "an appreciable effect" on the receiving waters. Before being amended by the Environment Act 1995, Schedule 10 to the WRA 1991 had conferred power on the National Rivers Authority to dispense with the requirement to advertise applications where it considered that the discharge would have "no appreciable effect" on the receiving waters. This was simply a re-enactment of the equivalent provision contained in Schedule 12 to the Water Act 1989, which in turn had retained the provision from section 36(1) of COPA 1974. Guidance on the exercise of this discretion was laid down in DoE Circular 17/84: the main test being that a change is not to be considered appreciable if there is less than a 10% increase on all relevant parameters, unless some significant environmental amenity is affected. These guidelines provoked considerable controversy.²⁸ In particular, it was argued that a 10% decline in water quality could in some cases constitute a "minor environmental disaster, whereas in other situations, where the quality of the receiving water is especially high, it will constitute an insignificant difference to the overall water quality."²⁹ Although the Environment Act 1995 removed the reference to the concept of an appreciable effect, its use in the 1996 Regulations makes clear that this concept still remains at the heart of the Environment Agency's discretion to dispense with the obligation to notify the public concerning an application for a discharge consent. In addition, the Regulations do not provide a statutory definition of the term, hence, as Howarth pointed out, an important publicity procedure is still left to hinge on "an avoidable degree of uncertainty".³⁰ Draft Regulations have not yet been prepared in this context for Northern Ireland; however it is likely that the DoE(NI) will follow the approach adopted by the Environment Agency in England and Wales.

A number of further points should be noted concerning the public notification requirements for applications for discharge consents. First, it is disappointing that Schedule 1 to the Water Order does not bring Northern Ireland into line with the requirements imposed on the Environment Agency to consult with relevant bodies outside the Agency concerning applications

²⁷ SI 1996/2971.

²⁸ Howarth, "Water Pollution: Improving the Legal Controls" (1989) 1 *Journal of Environmental Law* 25, at p 34.

²⁹ *Ibid.*, at p 35. See also: Eastwood & Ord, "Implementing Part II of the Control of Pollution 1974" (1986) *Water Pollution Control* 241; Matthews, "Part II of the Control of Pollution Act 1974: What it Means" (1987) *Water Pollution Control* 140.

³⁰ Note 28 *supra*.

for discharge consents.³¹ In the Northern Ireland context, the DoE(NI) should at least have been obliged to consult with the Northern Ireland Water Council, the Water Appeals Commission, the Council for Nature Conservation and the Countryside, the relevant fisheries board,³² the relevant district council(s) and specialist non-statutory interests.³³ Under the Water Resources Act 1991,³⁴ (WRA), the Agency's statutory consultees have six weeks within which to make representations concerning the application, and the Agency is required to consider any representations submitted within this period. Secondly, even where the DoE(NI) does require an applicant to advertise their application under Schedule 1, the Department, unlike the Environment Agency, is not placed under a statutory obligation to consider any representations made by members of the public in response to the notification. Not only is the Environment Agency under an obligation to consider public representations,³⁵ a statutory period of six weeks is laid down for public consultation. While the Secretary of State (in England and Wales) retains the power to change this period by regulation, the Water (NI) Order 1999 does not require the DoE(NI) to allow a specific period for public consultation. Finally, it should be noted that the public notification requirements concerning discharge consents issued without an application (under paragraph 4, Schedule 1 – discussed below) are more stringent than those applying in relation to consents issued on foot of an application. As is the case under Schedule 10 to the WRA 1991, the Department is obliged to publish notice of such a consent “in such a manner as may be prescribed”. Under the Control of Pollution (Applications, Appeals and Registers) Regulations 1996 (which govern the notification of such consents in England and Wales), the Environment Agency may only dispense with the requirement to advertise if the application is exempted from inclusion in the public register on the grounds of national security or commercial confidentiality.³⁶ The Environment Agency does not retain a discretion to dispense with the obligation to advertise in this context using the “appreciable effect” test.

Proposal No.2: The DoE(NI) should be empowered to require the Water Appeals Commission to hold a public inquiry, either alone or jointly with the Planning Appeals Commission, to consider objections to applications for discharge consents or where the proposed discharge is environmentally critical. The Department should be required to advertise its decision to hold a public inquiry.

³¹ Schedule 10, para 2. This requirement was introduced into the WRA 1991 by the Environment Act 1995. Regulation 5 of the Control of Pollution (Applications, Appeals and Registers) Regulations 1996 (SI 1996/2971) lists local authorities, other Ministers, local fisheries committees and the harbour authority.

³² The Foyle Fisheries Commission or the Fisheries Conservancy Board for Northern Ireland. See also note 12 *supra*.

³³ For example: the RSPB(NI), the Ulster Wildlife Trust, the National Trust, World Wide Fund for Nature (NI), Friends of the Earth (NI), the Northern Ireland Environmental Link, the Conservation Volunteers (NI).

³⁴ Schedule 10.

³⁵ Schedule 10, para.2(3).

³⁶ Schedule 1, para 1.

These proposals were only partially adopted by the new Water Order. Schedule 1 confers a new power on the Department to direct the Water Appeals Commission (WAC)³⁷ to conduct a public inquiry into applications for discharge consents.³⁸ Surprisingly, the Order does not confer a power to direct that the Planning Appeals Commission (PAC)³⁹ and the WAC conduct joint inquiries where a proposed development raises both pollution control and planning issues. Instead, paragraph 3(4) provides that where the DoE(NI) receives public representations in response to an advertised consent application which are relevant to a public inquiry to be held by the PAC, the Department may direct the PAC to consider those representations. However, in both instances, the DoE(NI) is obliged to consider the report on the public inquiry in making its decision whether to grant the discharge consent.

While the Department's power to direct that a public inquiry be held is undoubtedly welcome, public participation in the inquiry process is kept somewhat at arm's length despite the ethic underlying the concept of a public inquiry. To begin with, the Department is only obliged to notify the applicant for the consent and the relevant fisheries body⁴⁰ of a decision to direct the WAC to hold an inquiry or a direction to the PAC to consider representations made in response to the application.⁴¹ Even members of the public or environmental non-governmental organisations (NGOs) who have submitted representations to the Department concerning an advertised application for a discharge consent are not entitled to such notification. A further inconsistency with the concept of public participation lies in the fact that only the applicant and the relevant fisheries bodies are given a statutory right to appear and be heard at the inquiry. The Consultation Paper recommended that the right of appeal normally granted to an applicant and the fisheries bodies should be extinguished when a public inquiry is held as a second hearing of the issues would be inappropriate in such circumstances.⁴² To compensate for this loss, it was recommended that the fisheries bodies (but not the applicant) should be given a statutory right to appear and be heard at the inquiry. Article 13(4) does extinguish the right of appeal as suggested, but Schedule 1 paragraph 3(5) affords only the applicant and the fisheries bodies statutory rights to appear and be heard. Third parties are neither afforded a right to appeal a decision to grant a discharge consent nor a right to appear and be heard at a public inquiry conducted under the Water

³⁷ The Water Appeals Commission is an independent non-departmental body established under article 7 of the Water and Sewerage Services (NI) Order 1973 (SI 1973/70 (NI 2)).

³⁸ Para 3

³⁹ The Planning Appeals Commission is an independent non-departmental body established under the Planning (NI) Order 1972. The PAC now exists by virtue of article 110(1) of the Planning (NI) Order 1991. For further discussion of the powers and status of the PAC, see: Dowling, *Northern Ireland Planning Law*, Gill and Macmillan (1995), at pp 16-20.

⁴⁰ Either the Fisheries Conservancy Board for Northern Ireland or, if the discharge is to be made in the Derry area, the Foyle Fisheries Commission.

⁴¹ This is also the case under the Water Resources Act 1991. In this context the applicant may include the applicant for the discharge consent, and the applicant for planning permission, para 3(6).

⁴² Para 5.2.7.

(NI) Order 1999 - not even directly interested parties such as local residents or NGOs. Objectors in Great Britain are in a similarly weak position.

In this context it is also interesting to note that the Consultation Paper proposed that public inquiries should be held either to consider objections received in response to an advertised application or where the proposed discharge was considered to be “environmentally critical”. During their response to the Consultation Paper the Royal Society for the Protection of Birds (NI) (RSPB (NI)) described the phrase ‘environmentally critical’ as being fraught with potential inconsistencies and subjectivity. To ensure that such discharges were correctly and consistently identified, the RSPB(NI) urged the Department to implement integrated management plans for all major catchments in Northern Ireland. In addition, the RSPB(NI) recommended that there should be stronger representation of conservation NGOs in the composition of the Northern Ireland Water Council⁴³ to assist in the identification of such discharges and to facilitate further consideration of the conservation issues raised by each application. The Government’s response was to omit reference to this difficult phrase. Indeed, as is the case under the WRA 1991, no statutory pre-conditions are imposed on the Department’s power to call a public inquiry. More importantly, however, in the context of public participation is the Government’s refusal to include a representative from any of Northern Ireland’s environmental NGOs in the composition of the Water Council. Despite being a potent means of enhancing the quality of public debate on the environment and ultimately the quality of decision-making in this regard, no provision is made for representation of even the principal environmental interest groups operating within the Province.

Proposal No. 3: The DoE(NI) should be required to advertise appeals in relation to discharge consents and decisions to review discharge consents.

Both proposals were rejected. Neither the DoE(NI) nor the Environment Agency is required to advertise their intention to review a discharge consent, despite the fact that a review may lead to a significant change in the consented discharge. Although it is disappointing that the Government did not take the opportunity posed by the making of the Water Order to depart from this dubious position, what is more remarkable is that the Order does not bring the DoE(NI) into line with the position in England and Wales concerning public notification of appeals. Section 91(2B) of the Water Resources Act 1991 (WRA) provides that appeals against decisions concerning discharge consents shall be advertised if this is required by regulations adopted under the Act. In 1996 the Control of Pollution (Applications, Appeals and Registers) Regulations 1996⁴⁴ were adopted under these powers. Regulation 12(2) imposes an absolute obligation on the Secretary of State to publish notices of appeals within 21 days of the hearing.

⁴³ The Northern Ireland Water Council is now governed under article 58 of the 1999 Order. Its function is to advise the DoE(NI) and the DANI in relation to their functions under the 1999 Order and the DoE(NI) in relation to its functions under the Water and Sewerage Services (NI) Order 1973.

⁴⁴ SI 1996/ 2971

The new Water Order does not impose an equivalent obligation on the DoE(NI), nor does it confer power on the Department to adopt regulations that might require public notification of appeals against decisions concerning discharge consents. This anomaly is further compounded by the fact that third parties are completely excluded from participating in the appeals process. Under article 13 of the 1999 Order only the applicant or holder of the consent and the two fisheries bodies are conferred with rights to appeal the Department's decisions in relation to discharge consents. The Consultation Paper did not make any proposal to depart from this position. The lack of an obligation to provide public notification or powers to make such provision is also anomalous given that article 30 requires that all appeals in relation to discharge consents under article 13 must be recorded in the register established under the Order.

Proposal No. 4: The categories of information contained on public registers concerning the operation of the discharge consent should be expanded.

Consistent with Government and EU policy on public participation,⁴⁵ the Water (NI) Order 1999 considerably expands the range of information that must be recorded on the public register maintained by the DoE(NI) concerning the administration and enforcement of the Order. Indeed, it should be noted that, with the exception of one significant omission, the Order brings the Northern Ireland water pollution register fully into line with the equivalent requirements laid down in the Water Resources Act 1991⁴⁶ as amended by the Environment Act 1995.⁴⁷ The Order also brings the provisions governing access to the register into line with the Environmental Information Regulations (NI) 1993⁴⁸ which implement Directive 90/313/EEC⁴⁹ on freedom of access to environmental information. As currently drafted, the terms of the Freedom of Information Bill, which is discussed further below, will not directly affect the public's right to information held on statutory pollution registers such as that established under the Water Order. However, when it is enacted, Regulations will be adopted pursuant to the Freedom of Information Act for the purposes of implementing the requirements of the 1998 Århus Convention⁵⁰ in the United Kingdom. These Regulations will relate to the operation of pollution registers and will undoubtedly necessitate the amendment of the terms of the Water Order governing the pollution register. Finally, it should be noted

⁴⁵ Note 15 *supra*.

⁴⁶ Sections 190, 191A and 191B.

⁴⁷ Section 120, Schedule 22, para 170.

⁴⁸ SR 1993 No. 45.

⁴⁹ OJ L 158, 23 June 1990.

⁵⁰ UN Economic Commission for Europe *Convention on Access to Information, Public Participation in Decision-Making and Access to Justice in Environmental Matters*, signed at Århus, Denmark in June 1998; ECE/CEP/43. The full text of the Convention is available at www.eel.nl/treaties/CEP43E.htm. The "Århus Convention" (as it has become known) was signed by 39 European Governments (including the United Kingdom) and the EU itself. For general discussion of the key terms of the Århus Convention see ENDS Report 279, pp 43-44 and 292, pp 43-44. It should be noted that the adoption of the Århus Convention will lead to the adoption of an improved EC Directive on access to environmental information.

that given the confines of space and the fact that these provisions are not unique to water law, the new provisions governing Northern Ireland's water pollution register will not be the subject of detailed analysis in the present context.⁵¹

Article 30 of the Water Order requires the Department to maintain a public register that records virtually every aspect of the administration and enforcement of the Order. The modernised register will effectively enable environmental interest groups and individuals to obtain a three dimensional picture of how the core elements of the regulatory system are administered and enforced: namely, the discharge consent system, the system of water quality objectives and Integrated Pollution Control for prescribed processes. Not only will the public be able to identify the location and nature of consented discharges, but also the extent to which the Department is prepared to refuse its consent. In addition, the register will enable members of the public to monitor the Department's exercise of its enforcement and administrative powers under the Order – particularly the extent to which the DoE(NI) takes action to prevent pollution incidents and prosecutes offences under the Order. One notable omission are the discharge standards set for the DoE(NI)'s sewage and water treatment works. Despite the recommendation made by the *Review of the Water Act (NI) 1972: A Consultation Paper* that such information be included in the register, this proposal was not adopted by the Government. Consequently, this information remains available to the public only under the discretionary register⁵² established by the Department in 1996 and the general right to environmental information laid down in the Environmental Information Regulations (NI) 1993. It is also hoped that the DoE(NI) will exercise its powers under article 30 to adopt Regulations equivalent to those adopted by the Environment Agency⁵³ which prescribe when entries must be made and when information may be removed. The provisions governing the practical arrangements concerning access are also brought into line with the terms of the WRA 1991 and the requirement of the Environmental Information Regulations (Northern Ireland) 1993. Whereas the Water Act (NI) 1972 simply entitled the public to consult the register at all reasonable hours, article 30(3) of the Order introduces the additional requirements that consultation must be free of charge and reasonable facilities must be provided for obtaining copies of entries in the register – although a reasonable charge can be levied for copies made.

There are, however, a number of weaknesses in the water pollution register – all of which are shared with the equivalent register maintained under the Water Resources Act 1991. Neither the Environment Agency nor the DoE(NI) is required to record the fact that they have initiated a review of a

⁵¹ Bell, note 3 *supra*, pp 162-164.

⁵² The informal register was established as a result of commitments made to the House of Commons Environment Committee and targets set down in the *Environment Service Corporate Plan 1994-97*, at p 32. In relation to the commitments made to the House of Commons Committee, see the Minutes of Evidence submitted to the Committee by the Northern Ireland Office on July 19, 1993; HC 861-I, Session 1992-93, para 24.2.

⁵³ The Control of Pollution (Applications, Appeals and Registers) Regulations 1996, SI 1996/2971.

discharge consent. In light of the fact that public notification of reviews is not mandatory (as opposed to an application for a variation), this aspect of the regulatory function is essentially conducted in complete secrecy. In addition, neither the Water Order nor the WRA 1991 places the regulator under an obligation to record the fact that a public inquiry was conducted. Once again, given that public notification of an impending inquiry is not mandatory, this omission represents another important gap in the range of information provided to the public in this context. Neither the DoE(NI) nor the Environment Agency is required to record convictions for giving false information, or refusals to comply with directions requesting information concerning discharges to water. Another significant omission from both the Water Order and the WRA 1991 is that neither register must record the payment of compensation. Provision is made in several contexts for payment of compensation to potential or actual dischargers. It is submitted that information relating to the frequency and amounts of compensation paid would provide valuable information concerning the extent to which the taxpayer in the United Kingdom absorbs the costs of pollution control.

Important new controls have also been imposed on the exclusion of information from the water pollution register. Where previously the Water Act (NI) 1972 was silent as to the exclusion of information from the register, the Water Order explicitly addresses this important issue. As is the case under the WRA 1991,⁵⁴ information can be excluded from the register if it is necessary in the interests of national security or commercial confidentiality. However, where information is excluded, the DoE(NI), like the Environment Agency, must insert a statement into the register indicating the existence of the excluded information.⁵⁵ The strength of this requirement is unfortunately blunted by the fact that neither regulator is required to specify the grounds on which information is excluded. This weakness is however somewhat ameliorated by the wider rights and duties concerning public access to environmental information laid down in the Environmental Information Regulations (NI) 1993. Although the Regulations do not apply to environmental information that must be contained in public registers,⁵⁶ such information does not entirely escape their control. Regulation 6 provides that the practical arrangements made for the provision of environmental information, which must be released to the public under the terms of *other* statutory provisions, must be such as to secure conditions identical to those required under the Regulations. Regulation 4(2) provides that where a public body decides to refuse access to requested information, the reasons for the refusal must be provided. It is therefore arguable that a statement indicating the exclusion of information from the register maintained under the Water Order should always specify the reason on which the exclusion is based. Certainly, if the DoE(NI) were requested to provide an explanation for the exclusion and refused to do so, it would be acting in breach of regulation 6. Regrettably, members of the public are not granted a right of appeal against a decision of the DoE(NI) or the Secretary of State for Northern Ireland to

⁵⁴ Sections 191A and 191B inserted into the WRA 1991 by the Environment Act 1995.

⁵⁵ Article 30(2).

⁵⁶ Regulation 3(c).

exclude information from the water pollution register, nor is there an equivalent right of appeal against a refusal under the 1993 Regulations.

As already stated, the Freedom of Information Bill does not relate directly to pollution registers. Indeed, due to the considerably more generous provision for public access to environmental information under EC Directive 90/313/EEC on freedom of access to environmental information, the Government will have to disapply important sections of the Bill to ensure compliance with the Directive. However, Regulations will be adopted pursuant to the Act (when it is finally enacted) for the purposes of implementing the terms of the 1998 Århus Convention,⁵⁷ which will relate to the operation of pollution registers.⁵⁸ Article 5.2 of the Convention requires Governments to ensure that practical arrangements are made by public authorities to ensure that environmental information is made available to the public in a “transparent” and “effectively accessible” manner. Article 5 goes on to identify a number of specific means by which this objective may be achieved, in particular, by providing information contained in registers free of charge and ensuring that such information becomes progressively available through electronic databases which are easily accessible to the public through the Internet. Article 4 of the Convention governs the grounds for refusing access to environmental information. While Article 4 does not change the traditional grounds for exempting information from disclosure, it requires the balance of advantage to be shifted in favour of releasing information that would normally be exempted. Whereas Article 32 of the Water Order empowers the DoE(NI) to exempt information from disclosure on the grounds that it is commercially confidential if its release “would prejudice to an unreasonable degree the commercial interests of any individual or person”, Article 4.4(d) of the Convention considerably limits the type of information that could be exempted on these grounds. It provides that the safeguard for commercial confidentiality may only be used “where confidentiality is protected by law in order to protect a legitimate economic interest” but in any event, “information on emissions which is relevant for the protection of the environment shall be disclosed”. Volunteered information is only protected from disclosure under the Convention where it has been requested “without that party being under or capable of being put under a legal obligation” to pass it on to an authority.⁵⁹ In contrast, Article 32(4) of the Water Order currently enables the DoE(NI) to exempt information on the grounds that it is commercially confidential even where it has been furnished to the Department under a statutory duty. Finally, Article 4.4 imposes an overarching obligation which further shifts the balance of advantage in favour of disclosure. Article 4.4 concludes by providing that all grounds for refusing disclosure “shall be interpreted in a restrictive manner, taking into account the public interest served by disclosure and taking into account whether the information requested relates to emissions to the environment”. Although the Water Order empowers the DoE(NI) to disclose information notwithstanding that it is deemed to be commercially confidential where there is a public interest in doing so, there is no equivalent power to disclose information that would normally be exempted

⁵⁷ Note 50 *supra*.

⁵⁸ Draft Regulations will be published at the end of the year 2000.

⁵⁹ Article 4.4(g).

on the grounds of national security. Furthermore, the Order is silent as to the exemption of information concerning emissions to the environment. It is clear therefore that Article 32 of the Water Order will need to be amended to take account of the requirements of the Århus Convention.

Implementing the 'Polluter Pays' Principle

The 'polluter pays' principle encapsulates one of the most important shifts in contemporary environmental policy during the past decade.⁶⁰ Although originally popularised by the Organisation for Economic Co-Operation and Development in the early 1970s, the European Union is largely responsible for its integration into mainstream environmental policy in the United Kingdom. Because the environment was traditionally regarded as a 'free' economic resource, pollution has generally been perceived as having no economic cost. Consequently, although polluting emissions and discharges to the environment have been subject to regulation since the early nineteenth century, the taxpayer has absorbed the cost of administering the system of control. By the same token, penalties for pollution offences were low and the principal responsibility for clean-up lay with the regulator – and therefore the taxpayer. As consciousness of the true impact of pollution came to the political fore, this perception began to change. By the mid-1980s the European Commission had adopted the polluter pays principle as one of the key concepts underlying EC policy on the environment.⁶¹ In essence, the Community adopted the view that those persons whose activities give rise to actual or potential pollution should pay the costs of policing and remedial measures.⁶² During the same period, the Government under Mrs Thatcher expressed its wider political commitment to the ideology of privatisation and the development of market-based regulatory strategies. Although it was legally obliged to do so as a Member State of the EU, the Government, not surprisingly, readily embraced the polluter pays principle as part of the United Kingdom's environmental law and policy. In 1990, *This Common Inheritance*⁶³ stated that administrative regulation, working alone, was an

⁶⁰ For a discussion of the nature and role of this principle in contemporary environmental law and policy see generally: *This Common Inheritance*, note 2 *supra* at p 13 (para 1.28), p 136 (para 10.3) and p 168 (paras. 12.23-25); *Sustainable Development: The United Kingdom Strategy*, note 2 *supra*, Chapter 3, para 3.16; *A Better Quality of Life*, note 2 *supra*, at p 23; *Towards Sustainability*, note 3 *supra*, Chapter 7, section 7.4; Hughes, *Environmental Law*, (3rd ed 1996), at pp 21-23; Kramer, *E.C. Treaty and Environmental Law*, (2nd ed, 1995), at pp 56-57; Kramer, (1992) note 15 *supra*, Chapter 11; Jans, "Objectives and Principles of EC Environmental Law", in Winter, (ed.), *European Environmental Law: A Comparative Perspective*, Dartmouth (1996), at pp 285-287; Kiss & Shelton, *Manual of European Environmental Law*, (2nd ed, 1997), at p 43; Alder & Wilkenson, note 15 *supra*, at pp 171-184; Sands, *Principles of International Environmental Law: Volume 1*, Manchester University Press (1995), at pp 213-217; O'Riordan, *Ecotaxation*, Earthscan (1997).

⁶¹ See: Kramer, (1992) note 15 *supra*, Kramer, (1995) note 60 *supra*, and *Towards Sustainability*, note 3 *supra*. It should also be noted that the 'polluter pays principle' was enshrined in article 130R of the EC Treaty in 1992 as a result of amendments made by the Treaty on European Union. Article 130R has recently been renumbered as article 174 EC under the Treaty of Amsterdam.

⁶² Kramer, *ibid.*

⁶³ Note 2 *supra*, and at p 13, para 1.26.

expensive and ineffective means of controlling environmental pollution. Although the Government maintained that regulation would remain at the heart of Britain's system of environmental control for the foreseeable future,⁶⁴ *This Common Inheritance* emphasised the Government's commitment to developing new market-based instruments for environmental protection. In particular, *This Common Inheritance* made it clear that the polluter pays principle would underlie all forms of environmental controls in the United Kingdom, including the system of administrative regulation.⁶⁵

The polluter pays principle was integrated into the system of regulation governing water pollution in England and Wales in three principal ways: namely through increasing the maximum penalties for water pollution, strengthening polluters' liabilities for pollution prevention and clean-up costs and introducing a system of discharge consent fees. The Water Act 1989 introduced powers for the establishment of a system of discharge consent fees; these powers were exercised almost immediately and a system came into force in July 1991.⁶⁶ Penalties for water pollution offences were increased significantly by the Environmental Protection Act 1990.⁶⁷ Prospective polluter's liabilities for pollution prevention and polluter's liabilities for clean-up costs have been gradually increased in England and Wales through successive reforms, particularly under the Water Act 1989 and the Environment Act 1995.⁶⁸ As usual, several years elapsed before equivalent measures were introduced in Northern Ireland. The first steps to implement the polluter pays principle in Northern Ireland were taken relatively quickly; however, this momentum evaporated as the programme of reform turned to focus almost exclusively on the implementation of EC Directives on the environment. Three years passed before equivalent penalties for water pollution offences were introduced into Northern Ireland in 1993.⁶⁹ A decade passed before the introduction of equivalent powers to introduce discharge consent fees – now contained in article 11 of the Water (NI) Order 1999. The Water Order also finally introduces the enhanced liabilities for pollution prevention and remediation introduced in England and Wales in 1989 and the subsequent amendments to the Water Resources

⁶⁴ *Ibid.*

⁶⁵ *Ibid.*, and at Annex A.

⁶⁶ These powers were originally introduced under Schedule 12, para 9 of the 1989 Act. They were subsequently consolidated in section 131 of the Water Resources Act 1991 and are now contained in the general charging provisions of sections 41-42 of the Environment Act 1995. Equivalent charges are being introduced in Northern Ireland in respect of licences granted under the Waste and Contaminated Land (NI) Order 1997 (SI 1997/2778 (NI 19)) and the Industrial Pollution Control (NI) Order 1997 (SI 1997/2777 (NI 18)).

⁶⁷ c.43.

⁶⁸ In this regard, see the discussion of anti-pollution works notices in Part One of this article, (2000) 51 *NILQ* 65 *et seq.*

⁶⁹ Introduced under article 17 of the Water and Sewerage Services (Amendment) (NI) Order 1993 (SI 1993/3165 (NI 16)). These penalty provisions have now been consolidated in the Water (NI) Order 1999. However, in *Control of River Pollution in Northern Ireland*; HC 693, Session 1997-98, at p 81, the Comptroller and Auditor General for Northern Ireland expressed his concern that courts in Northern Ireland were not making use of the increase in the penalty ceiling for water pollution offences. See also Part One of this analysis, (2000) 51 *NILQ* 65 *et seq.*

Act 1991 made by the Environment Act 1995. The Department's new powers to impose prevention and clean-up obligations were addressed in Part One of this article;⁷⁰ however, the present section will focus on the nature and scope of its powers to establish a system of charges for discharge consent fees.

Article 11 of the 1999 Order provides that the DoE(NI) may introduce a scheme of charges for discharge consents in Northern Ireland.⁷¹ Charges may be imposed on applications for new consents, applications for variations of existing consents, retrospective consents granted without an application and consents granted under the Water Act (NI) 1972.⁷² Either the applicant or consent holder will be liable for payment of the fee. Failure to do so will entitle, but not oblige the Department to revoke the consent. As is the case in England and Wales, cost recovery is the essential philosophy underlying the charging scheme. Article 11(5) provides that the charging scheme must be sufficient to ensure, in so far as it is reasonably practicable, that the costs incurred by the Department in exercising its powers under the Order in relation to discharge consents are covered. Consequently consent fees will cover not only the costs incurred in making the initial decision to grant a consent, but also the Department's expenditure in monitoring the discharge during the life of the consent.⁷³ Although environmental interest groups broadly welcomed the introduction of discharge consent fees in Northern Ireland, the Royal Society for the Protection of Birds (NI) expressed their concern during the consultation process that the Department's charging powers were constrained by cost recovery. The RSPB(NI) voiced its support for the wider use of economic instruments in environmental policy and in particular called for the amendment of the proposed draft Order to allow for the introduction of incentive charging for water pollution. In this regard the Society pointed to the Government's Statement of Intent on Environmental Taxation⁷⁴ and to the proposals for incentive charging set out by the Department of the Environment, Transport and the Regions' 1998 consultation on *Economic Instruments for Water Pollution*.⁷⁵ Regrettably, this proposal was not adopted.

⁷⁰ *Ibid*, at pp.20-31.

⁷¹ This system will not apply to effluent from domestic septic tanks serving single dwellings.

⁷² Article 11(1). Article 11(6)-(8) imposes strict public notification requirements in relation to proposals to introduce discharge consent fees. The DoE(NI) must publish notices of a proposal to introduce a scheme of charges which specify the period within which representations may be made and must consider any representations duly made. Unlike the public consultation requirements relating to applications for discharge consents or variations thereof, the Department must bring the provisions of any charging scheme to the attention of persons likely to be affected by it. The Department does not have a residual power to dispense with this requirement.

⁷³ In this regard, the fee will cover the costs incurred in sampling discharges, inspecting discharges, discharge related impact monitoring, work on the review of consents, laboratory services and direct administration connected to these matters.

⁷⁴ HM Treasury (1997), *Environmental Taxation: Statement of Intent*; HM Treasury Press Office, 2 July, 1997.

⁷⁵ Consultation Paper, DETR (1997).

Although the DoE(NI) is not obliged to establish a scheme of charges, the Government's commitment to the polluter pays principle is such that the introduction of discharge consent fees in Northern Ireland is inevitable. The DoE(NI) has not yet published detailed proposals for a charging scheme; however, the *Compliance Cost Assessment* accompanying the proposal for a draft Water (NI) Order noted the Department's intention to adopt the approach used by the Environment Agency in calculating charges in England and Wales. Consent charges in England and Wales are set according to a formula comprised of three separate elements, namely the volume of effluent produced, its content and the characteristics of the receiving water. Weighted bands have been established for each of these three elements; for example, in relation to volume, the greater the volume of effluent being discharged, the greater the weighting accorded to that element. The weightings accorded to each element of the discharge are multiplied together to give a final figure which in turn is multiplied by a national financial factor. By this means, the consent fee not only shifts the cost of regulation from the taxpayer to the polluter, but also provides the polluter with an incentive to reduce the volume or improve the quality of effluent being discharged as this would reduce the annual monitoring charge. Uniformity of charging is one of key principles underlying the charging system operating in England and Wales. The continuation of this pattern of charging consistency is likely to be fundamental to the debate surrounding the introduction of consent fees in Northern Ireland. However, in light of the fact that fees for Integrated Pollution Control authorisations⁷⁶ in Northern Ireland are already less expensive than the equivalent fees imposed in England and Wales, it is certainly possible that fees charged for consents granted under the Water Order will also be less expensive than the fees imposed for consents granted under the WRA 1991. However, cheaper consent fees will also raise the spectre of Northern Ireland becoming a more attractive location for polluting industries than elsewhere in the United Kingdom.

One final, but significant, problem surrounding the implementation of discharge consent fees in Northern Ireland concerns the practical problems faced by the DoE(NI)'s Environment and Heritage Service (EHS) in making the necessary arrangements to introduce the charging system. Although work began on the development of a cost recovery system in 1996, the Comptroller and Auditor General's 1998 report concerning the operation of the EHS made it clear that much work remained to be done before a system of discharge consent fees could be introduced in Northern Ireland.⁷⁷ The Auditor General report revealed that the EHS has no written policies or procedures or guidance either for its own staff or for its agents on the recovery of costs arising from pollution incidents.⁷⁸ The EHS told the

⁷⁶ Integrated Pollution Control authorisations are granted under the Industrial Pollution Control (NI) Order 1997.

⁷⁷ *Control of River Pollution in Northern Ireland*; HC 693, Session 1997-98, Part 8, pp 77-81.

⁷⁸ *Ibid*, p 79, para 8.8. In the absence of guidance on costs, the Auditor General examined what occurred in practice and reported the following (pp 79-80, para 8.10): The costs incurred by the EHS or its agents in investigating the sources of pollution and taking samples are neither recorded nor recovered. Where a successful prosecution is taken, the costs of analysis are recovered through the courts, but only that proportion of such costs as are necessary to establish the case

Auditor General that it was planning to draw up operational guidelines for dealing with incident costs when the new Water Order became law.⁷⁹ While the Auditor General recommended that work should progress on the preparation of an accurate assessment of the full costs of all consent work, the development of a charging strategy consistent with the policy objective of full cost recovery, and the development of a system to enable management to monitor the extent to which costs are being recovered,⁸⁰ the Auditor General's Report regrettably gave no indication as to how far work in this regard had progressed. Taking 1992 as the original target date for the enactment of the Water Order, the Auditor General estimated that by April 1998, the delays in conferring on the DoE(NI) the powers necessary to introduce a charging system had already cost the taxpayer up to £5 million.⁸¹ In light of the under-developed nature of the EHS' cost recovery system, it is unlikely that this central aspect of the polluter pays principle will be implemented in Northern Ireland in the short to medium term.

'Better Regulation' of the Discharge Consent System

One of the major policies underlying the reform of the discharge consent system is the Labour Government's policy of "better regulation".⁸² Like the Conservative Government's policy of deregulation,⁸³ better regulation is aimed at streamlining and simplifying regulatory controls; however, it also seeks to refine the system of environmental regulation to better reflect sustainable development objectives. In February 1998 the Government published *Opportunities for Change*⁸⁴ – a consultation paper setting out its vision of sustainable development and proposals for revising the United Kingdom's strategy for achieving sustainability in this country. *Opportunities for Change* reiterated previous national policy on sustainable

for prosecution. In relation to clean-up costs, only the cost of materials and hire of equipment used and the staff costs of certain agents employed by the EHS are recovered from polluters. However, EHS staff time is neither recorded nor recovered. In relation to enforcement action, legal costs are recovered if the prosecution is taken by one of the fisheries bodies – private sector legal services are used and their costs are recovered through the courts. However, if the EHS prosecutes through the Water Act (NI) 1972, legal services are arranged *via* the DPP and no costs are recovered. In the event of a successful prosecution, only those costs (except clean-up costs) which can be recovered direct from the polluter are recoverable through the courts.

⁷⁹ *Ibid*, p 79, para 8.9.

⁸⁰ *Ibid*, p 78, para 8.6.

⁸¹ *Ibid*, p 78 at para 8.6. The Auditor General did, however, note that the figure may be less if the EHS had phased consent fees in gradually as occurred in England and Wales.

⁸² *Explanatory Document for a Proposal for a Draft Water (NI) Order* (Published jointly by the DoE(NI) and DANI), at p 2. For further guidance on the scope of this concept, see: *The Better Regulation Guide*, Cabinet Office, Regulatory Impact Unit, August 1998; and *Modernising Government*, (1999) Cm 4310.

⁸³ In this regard, see: *Proposed Amendments to the Water Resources Act 1991 and the Reservoirs Act 1975: Joint Consultation Paper* (Water Directorate, DoE(GB) and the Environment Division of the Welsh Office: October, 1993).

⁸⁴ *Sustainable Development: Opportunities for Change – A Consultation Paper on a Revised United Kingdom Strategy* DETR 1998, <http://www.environment.detr.gov.UnitedKingdom/sustainableconsult1/inde.htm>.

development to the effect that regulation would continue to be essential to the achievement of sustainability, particularly in setting standards which protect human health and the natural environment.⁸⁵ However, this document also introduced the objective of “better regulation” – defined as regulation which “delivers objectives without imposing unnecessary burdens.”⁸⁶ The Government stated that “where regulations exist, we need to consider whether they are too narrowly focused on a single objective and ignore wider economic, environmental and social goals”.⁸⁷ In May 1999 the Government published *A Better Quality of Life: A Strategy for Sustainable Development in the United Kingdom*⁸⁸ which reiterated its commitment to ensuring that where regulation is used, it will comply with the Government’s principles of better regulation. These principles were expressed as requiring regulation that is “targeted at the problem in hand; clear and simple to understand; applied consistently, proportionate to the problem and the circumstances of individual businesses, voluntary groups and others; enforced effectively and constructively by the body accountable for its conduct”.⁸⁹

The policy of better regulation has informed a raft of new provisions in the context of the discharge consent system. Of particular relevance in this regard are the new provisions governing the application procedure for consents and the Department’s considerably expanded powers to monitor and review consented discharges. Some of these provisions continue the contemporary policy emphasis on inclusiveness and greater environmental protection by ensuring greater public consultation and the imposition of more stringent controls on consent holders, while others take account of the wider economic needs of consent holders, particularly the need for stability and clarity. However, the consistent impact of this policy is to streamline and simplify the controls imposed on the application procedure and the review powers which considerably clarifies the legal position of the applicant, the consent holder and the DoE(NI). Each of these changes will now be addressed in turn.

(i) New Provisions Governing the Application Procedure

One of the most obvious improvements introduced by the Water Order is the decision to place all the specific provisions governing applications for discharge consents and the procedures relating to review, revocation, modification and appeals and so forth into a schedule to the Order.⁹⁰ As is the case under the Water Resources Act 1991, which governs the consent system in England and Wales, this practice considerably clarifies the text of the legislation. However, the changes do not stop there. To begin with, all applications for discharge consent must now be made on the form provided by the DoE(NI).⁹¹ Where an application (made in compliance with the

⁸⁵ *Sustainable Development: The United Kingdom Strategy*, note 2 *supra*, at Chapter 33, pp 213-214.

⁸⁶ *Opportunities for Change*, note 84 *supra*, at para 85.

⁸⁷ *Ibid.*

⁸⁸ (1999) Cm 4345.

⁸⁹ P 27, para 5.10.

⁹⁰ Schedule 1.

⁹¹ Schedule 1, para 1(1)(a). This requirements also applies to applications for variations made by the holder of the consent – para 7(1).

requirements laid down in the Order)⁹² is received by the Department, it is now under a statutory duty to consider whether to grant or refuse the consent, or grant it subject to conditions.⁹³ In recognition of the increased complexity of the legislative framework governing water pollution and water quality, the Order extends the period within which the Department must make its decision from three to four months.⁹⁴ Not before time, the Order also reverses the default position in the event that a decision is not made within four months. Whereas failure in England and Wales to make a decision on an application for a consent within the statutory period has, for decades, meant that the application is deemed to be refused,⁹⁵ in Northern Ireland, the reverse has been the case. Under section 9(5) of the Water Act (NI) 1972 Act, if the Department failed to make a decision on an application within the statutory period of three months, an unconditional discharge consent was granted by default. Fortunately, the Water Order reverses this arcane emphasis on the primacy of economic interests. Hence, in the absence of a decision by the DoE(NI) within the statutory period, an application will be deemed to have been refused. It is, however, interesting to note that although the *Review of the Water Act (NI) 1972* recommended extending the decision-making period from three to four months,⁹⁶ it suggested that the application “would take its normal course” in the event that no decision was taken within the statutory time limit.⁹⁷ One can only wonder that the DoE(NI) was prepared to persist with this clearly anomalous approach to pollution control as recently as 1993. One further environmental safeguard introduced by the Order in this context is the removal of the applicant’s immunity from prosecution for discharges made pending the outcome of an application for consent. Under section 9(11) of the 1972 Act a person was not guilty of a water pollution offence if they had made an application for a discharge consent and the discharge made complied with the description of the discharge contained in the application. This immunity from prosecution lasted for the three months allotted to the Department to consider the application.

The Order also increases and clarifies the DoE(NI)’s powers to obtain the information necessary to determine the application. As was the case under section 9(1) of the 1972 Act, applicants must provide such information as

⁹² Or any requirements laid down in regulations adopted by the DoE(NI) under Schedule 1, para 1.

⁹³ Schedule 2(1).

⁹⁴ Thereby bringing Northern Ireland into line with the position in England and Wales (Schedule 10, para 3(2) WRA 1991). The 4-month time limit also applies to deciding applications to vary a consent (para 7). As is the case under the WRA 1991, the DoE(NI) retains the power to extend this period either by regulation or with the (written) agreement of individual applicants. In the event that a public inquiry is directed (either jointly with PAC or not), the time limit for making a decision on a discharge consent application will cease to apply. This reform implements recommendation 5.3 of the *Review of the Water Act (NI) 1972: A Consultation Paper*, December 1993 (DoE(NI) and Department of Agriculture (NI)).

⁹⁵ This has been the consistent position in England and Wales since the adoption of COPA 1974.

⁹⁶ Note 94 *supra*, at para 5.3.2.

⁹⁷ *Ibid.*

may be required by the DoE(NI) either on, or together with, their application form.⁹⁸ However the Department is also given a new power to serve a notice on the applicant requiring the provision of “further information of any description” that may be necessary for the purposes of determining the application. In addition, the penalty for knowingly or recklessly providing information that is false in a material particular for the purpose of obtaining a discharge consent is increased from £500 to the statutory maximum of £5,000 for summary conviction;⁹⁹ the alternative penalty of three months imprisonment is retained. The penalty for conviction on indictment remains at an unlimited fine and/or a maximum sentence of 2 years. It should be noted that while failure to provide the necessary information to support an application will not invalidate an application,¹⁰⁰ failure to comply with advertising requirements or requirements concerning the provision of information, entitles the DoE(NI) to refuse to proceed with the application either entirely or until the relevant requirements have been complied with. Schedule 1 also makes it clear that an application to discharge at two or more places may be treated as separate applications for each discharge.¹⁰¹

One important problem that emerged in the operation of the discharge consent system under the Control of Pollution Act 1974 (COPA) and Water Act (NI) 1972 concerned the absence of control over discharges of new pollutants. New pollutants included substances that were either added to the discharge after the consent had been obtained, or were incapable of being traced when the consent was issued, or were identified as having a polluting effect after the consent was issued. William Howarth pointed out that controlling new pollutants under COPA 1974 was extremely difficult due to the proliferation of new industrial chemicals and the absence of a general duty imposed on consent holders to inform the regulator of changes in the composition of the discharge.¹⁰² Even where new pollutants were identified, regulators were very restricted in their opportunities to review discharge consents.¹⁰³ Under COPA a Regional Water Authority could not review a discharge consent within two years of it being granted without the agreement of the discharger. In Northern Ireland, the Water (NI) Act 1972 afforded the holders of discharge consents a further year of immunity from review. Consequently, provided they complied with the conditions of their consents, unco-operative dischargers could continue to emit pollutants not covered by their consent for several years. This loophole was closed in England and Wales under the Water Resources Act 1991 and eight years later in Northern Ireland by the Water (NI) Order 1999. Both measures provide that it is not an offence to discharge or deposit any effluent or matter made “under and in accordance with” a discharge consent. Similarly it is not an offence to discharge or deposit any effluent or matter as a result of any act or omission “under and in accordance with” a discharge consent.¹⁰⁴ In legislative terms,

⁹⁸ Schedule 1, para 1(3).

⁹⁹ Article 29.

¹⁰⁰ Para 1(4).

¹⁰¹ Schedule 1, para 1(5).

¹⁰² Howarth, “Water Pollution: Improving the Legal Controls” (1989) 1 *Journal of Environmental Law* at pp 28-29.

¹⁰³ *Ibid.*

¹⁰⁴ Section 88 of the WRA 1991 and article 10 of the Water (NI) Order 1999.

this amendment goes some way to conveying the message that a discharge consent is not a licence to pollute. To clarify the matter further, the Environment Agency has recommended that all consents should include a general condition that excludes the discharge of any substance not specifically identified in the discharge consent. It is hoped that the DoE(NI) will adopt a similar practice in Northern Ireland. There is, however, a remaining weakness in the controls available in relation to new pollutants. This weakness is present in the WRA 1991 and has been carried over into the terms of the Northern Ireland Order. Although a discharge must be made “under and in accordance with” the terms of the consent, the Department does not have the power to vary a consent solely on the grounds that a discharge is being made in breach of the consent,¹⁰⁵ nor does it have the power to vary the consent for the purposes of addressing the presence of a new pollutant.¹⁰⁶ Unless the new pollutant is caught under new EC or international law, the Department cannot review a consent for at least four years. Alternatively, the Department could attempt to vary the consent within this period on the grounds of public health or protection of flora and fauna dependent on an aquatic environment. However, as will be discussed below, strict rules govern the Department’s powers in this regard and there is a potential liability to compensate the holder of the consent for loss incurred as a result.

(ii) Consents Granted Without Applications

The DoE(NI)’s powers to regularise illegal discharges have also been extended to come into line with the equivalent powers conferred on the Environment Agency.¹⁰⁷ Under the 1972 Act the Department’s powers in this respect were limited to granting a consent where an unlawful discharge had already occurred. However, Schedule 1, paragraph 4(1) of the Order also enables the Department to act in a preventative manner by extending this power to include situations where an unlawful discharge is likely to occur.¹⁰⁸ As was the case under the 1972 Act, polluters are not granted immunity for offences committed before the consent is granted under paragraph 4.¹⁰⁹ A particularly interesting feature of the Department’s powers in this regard concerns the requirements for public consultation. Whereas the 1972 Act was silent as to whether notice of a consent issued without an application should be published, the 1999 Order imposes very specific and stringent requirements concerning public consultation on the granting of such consent.¹¹⁰ Indeed, these requirements provide consultation guarantees that are not afforded to the public where the DoE(NI) receives an application for a discharge consent. The Department must publish notice of a consent issued under paragraph 4. There is no provision made for dispensing with this

¹⁰⁵ Bell, note 3 *supra* at p 460.

¹⁰⁶ *Ibid.*

¹⁰⁷ These are contained in Schedule 10, para 6 to the WRA 1991.

¹⁰⁸ Where the DoE(NI) grants a consent under paragraph 4, notice of this decision must be served on the relevant fisheries body (para 4(2)). It should also be noted that the DoE(NI) retains its powers to attach appropriate conditions to a consent issued under paragraph 4.

¹⁰⁹ Para 4(3).

¹¹⁰ The public consultation requirements are identical to those applying in England and Wales under Schedule 10, para 6 to the WRA 1991.

requirement and, consistent with the polluter pays principle, the cost of advertising the consent can be recovered from the polluter.¹¹¹ In addition, the Department must send copies of the consent to such persons as may be prescribed.¹¹² Furthermore, paragraph 4 places the Department under a statutory duty to consider any representations made in relation to this consent and confers powers to prescribe a statutory period within which such representations should be made. Under normal circumstances, once a discharge consent has been granted it is immune from revocation or review for at least four years unless with the holder's consent. However, paragraph 6(4) provides that this restriction will not apply to a consent issued under paragraph 4. If the Department considers it appropriate to modify or revoke the consent in light of representations received, and notice of this decision is served within three months of the period to be prescribed for making representations in this context, the consent may be revoked or modified. Consequently, although public consultation takes place after the consent has been granted, it is a meaningful process. The effectiveness of public consultation in relation to paragraph 4 consents depends ultimately on the DoE(NI)'s ability to ensure that a notice of its intention to alter or revoke the consent is issued within the three month time limit. Whereas failure on the part of the DoE(NI) to make a decision in relation to an application for a discharge consent simply means that the application is refused, failure to issue a notice reviewing or revoking a paragraph 4 consent effectively renders it immune from review for at least four years.¹¹³ Hence, on the one hand, the Order denies the regulator the benefit of the extended period given for considering applications for discharge consents, while on the other hand it permits polluters, who have failed to apply for a consent and may be guilty of a pollution offence, to rely on the extended period of immunity from review granted to those who have applied for their consent before making a discharge. Given that the issues surrounding a discharge made under a paragraph 4 consent may be no less complex than those made under normal consents, this imbalance is unfortunate to say the least.

(iii) The Power to Review a Discharge Consent

As was the case under the Water Act (NI) 1972, the Water Order confers powers on the DoE(NI) to review discharge consents periodically; however, under the new legislation these powers are considerably strengthened. The DoE(NI)'s powers to review discharge consents are now identical to those available to the Environment Agency under the WRA 1991. The most important change in the Department's power to review is that discharge consents are no longer immune from revocation. Under the 1972 Act, provided a consent was being used (that is, a discharge was being made), a consent could only be varied.¹¹⁴ Although it could restrict the nature or scale

¹¹¹ Schedule 1, para 4(7).

¹¹² Para 4(5).

¹¹³ Subject to the DoE(NI)'s general powers to review at any time for the purposes of ensuring compliance with EC and international law, or to protect public health, or flora and fauna dependant on an aquatic environment. The Department's powers in this respect are discussed below. It should be further noted that, as is the case for consents issued in response to an application, the holder of a para 4 consent may apply at any time for a variation of the consent – para 6(5).

¹¹⁴ Section 10(2).

of a discharge by varying the conditions attached to the consent or by imposing new conditions, the Department did not have the power to revoke the consent entirely. Not before time, the 1999 Order remedies this fundamental weakness in the regulatory system. Schedule 1, paragraph 5 of the Order now gives the Department power to conduct periodic reviews that may result in the revocation or modification of a consent, or the imposition of conditions on a previously unconditional consent.¹¹⁵ Although recommended in the Consultation Paper,¹¹⁶ the Order does not oblige the Department to advertise its intention to conduct a review. Given that a review may result in a significant change to the consented discharge, and in light of the general thrust towards enhancing public participation in the regulation of water pollution, this omission is an important weakness in the legislation. Public notification of a review will only be required where the discharger actually applies for a variation of the consent. In such situations, the notification and consultation requirements governing an initial application for a consent will apply.¹¹⁷ The weakness in public consultation is further compounded by the fact that the Order does not explicitly require the Department to record the conduct and outcome of reviews on the public register maintained under article 30.

As was the case under the 1972 Act, the Department's power to conduct a review is entirely discretionary. There are however, some important limitations imposed on the exercise of this power. Prior to the enactment of the Water Order, consents were immune from modification for at least three years - in contrast to dischargers in England and Wales who were, until 1996, only allowed a two-year period of immunity from review.¹¹⁸ In Northern Ireland, as in England and Wales, this period has now been extended to four years.¹¹⁹ Although some would argue that reviews should be carried out more rather than less frequently, the Consultation Paper recommended this change in the interests of providing greater stability for dischargers and therefore more security in investment and planning. It is at least welcome that dischargers in Northern Ireland are no longer afforded a longer period of immunity than their counterparts in England and Wales. In the interests of better regulation, every discharge consent must specify the period during which the consent cannot be modified or subject to new conditions without the agreement of the discharger.¹²⁰ One interesting exception to the four-year

¹¹⁵ The Department's power in this regard are identical to those conferred on the Environment Agency under Schedule 10, para 7(2) of the WRA 1991.

¹¹⁶ Note 94 *supra*, at para 5.5.1.

¹¹⁷ Schedule 1, para 7(2).

¹¹⁸ This was the original position under COPA 1974, which persisted until recently under the Schedule 10, para 7(3) of the WRA 1991.

¹¹⁹ Schedule 1, para 6(3) of the Water (NI) Order 1999. In England and Wales the statutory period of immunity was increased to four years by the Environment Act 1995 (section 120, Schedule 22, para 183), which came into effect in 1996 under the Environmental Act 1995 (Commencement No.8 and Saving Provisions) Order 1996 SI 1996/2909.

¹²⁰ Consent holders now have a new right under article 13(1)(e) to appeal a decision by the Department to specify a period of less than four years without the holder's consent. The requirement concerning the specification of the review period also applies to a consent that has been modified as a result of a previous review - Schedule. 1, para 6(2). It is interesting to note, however, that the Department is

rule concerns the Department's power to revoke a consent. Although the DoE(NI) is not conferred with the general power conferred on the Secretary of State in England and Wales to direct that a review be undertaken at any time, the Department's power to revoke a consent does not appear to be subject to the four-year rule.¹²¹ The Order also shortens the potential 'life' of a discharge consent. Under the 1972 Act, a discharge consent had a minimum life of three years; however, the DoE(NI) now has the right to revoke any consent that has not been used at any time during the previous year. Given its general power to revoke, this power is somewhat superfluous.

In addition to its general power to review discharge consents every four years, Schedule 1, paragraph 5(4) gives the Department new powers to review consents *at any time* in specific circumstances. Paragraph 5(4) provides that the DoE(NI) may review a consent (leading to revocation, modification or the imposition of conditions) at any time, if it appears that it is necessary either: (a) to implement any of the United Kingdom's EC or international law obligations or (b) to protect public health or flora and fauna that is dependent on an aquatic environment.¹²² Before proceeding to address the nature and scope of these new powers, it is important to emphasise, that the DoE(NI), like the Environment Agency, is not given power to vary a consent prematurely purely on the grounds that the discharge does not comply with the terms of the consent.¹²³

Although the DoE(NI)'s new power to review a consent at any time to ensure compliance with EC and international law is undoubtedly welcome, it is disappointing, particularly in relation to compliance with EC environmental law, that the decision to review for these purposes is left to the DoE(NI)'s discretion – particularly given the Department's poor record in implementing EC Directives on the environment.¹²⁴ It is likely, however, that the Northern Ireland Act 1998¹²⁵ – if revived – would remedy this weakness. Section 24(1)(b) of the Northern Ireland Act provides that "a Minister or Northern Ireland department has no power to make, confirm or approve any subordinate legislation, or *to do any act*" which is incompatible with EC law. Although section 24 does not define the meaning of an 'act' in this context or explain whether a failure to act would come within its ambit, it is certainly arguable that a failure to act that is incompatible with EC law would amount to a breach of section 24(1)(b). In effect, if EC law prohibits certain types of discharges to freshwater bodies, and the DoE(NI) has issued a consent that allows an illegal discharge to be made, then it can be argued that a failure on the part of the DoE(NI) to review the consent to bring it into line with EC law, would amount to an act that is incompatible with EC law under section

not obliged to provide the holder with specific notification of its intention to review after than period has expired.

¹²¹ Schedule 1, para 6(1) and (2).

¹²² This power is identical to the power given to the Environment Agency under Schedule 10, paragraph 7 of the WRA 1991.

¹²³ Bell, note 3 *supra* at p 460.

¹²⁴ S. Turner & K. Morrow, "The Impact of EU Law on the Environmental Law of Northern Ireland", in Holder (ed) *The Impact of EU Environmental Law in the United Kingdom*, Wiley (1997), pp 69-84.

¹²⁵ c.47. The terms of the 1998 Act were suspended under the Northern Ireland Act 2000, see *infra*.

24(1)(b). In addition, legal developments at both national and EC level are likely to provide strong incentives to the DoE(NI) to ensure practical compliance with EC law. In 1992 the Maastricht Treaty amended the EC Treaty to confer power on the European Commission to recommend to the European Court of Justice the imposition of financial penalties on Member States who fail to comply with a Court judgment declaring that they are in breach of their Community obligations.¹²⁶ The Commission declared its intention in 1996 to focus its enforcement activities on the specific issue of ensuring Member States' compliance with EC environmental law, monitoring in particular their compliance in practice as opposed to purely formal implementation.¹²⁷ In addition, the past decade has witnessed the European Court's rapid development of a principle of state liability under which Member States may be held liable to compensate private parties who have sustained injury or loss as a result of a State's failure to comply with its obligations under EC law.¹²⁸ Hence, non-compliance with EC environmental law is not only likely to attract enforcement action from the European Commission, Member States also face the possibility of simultaneous actions in damages before their national courts if loss or injury is sustained by private parties as a result of their failure. The Concordat on Co-Ordination of European Union Policy Issues¹²⁹ governing relations between the United Kingdom Government, the Scottish Parliament and the National Assembly for Wales states that financial costs and penalties imposed on the United Kingdom arising from a failure of implementation or enforcement of EC law by a devolved administration (on a matter falling within its responsibility) will be paid by the devolved administration. While an equivalent Concordat has not yet been published for the Northern Ireland Assembly, it is likely that similar arrangements would be agreed following the restoration of devolution. Given that responsibility for almost all matters concerning the environment are transferred to the Northern Ireland Assembly under the Northern Ireland Act 1998, the revival of devolution would mean that the financial impact of non-compliance with EC environmental law would be felt more immediately by the DoE(NI) than has previously been the case.

The Department's new powers to review a consent at any time in the interests of protecting public health or flora and fauna dependent on an

¹²⁶ Article 171EC (now renumbered article 228 EC under the Treaty of Amsterdam).

¹²⁷ *Implementing Community Environmental Law* COM(96) 500 final. For further discussion of the European Commission's use of its enforcement powers in the context of environmental law, see: Macrory and Purdy, "The Enforcement of EC Environmental Law Against Member States" in Holder (ed.), note 124 *supra*, pp 27-50; Macrory, "The Enforcement of Community Environmental Laws: Some Critical Issues" (1992) 29 *CML Rev* pp 347-369

¹²⁸ The seminal judgments on the principle of state liability under EC law were delivered in Case C-6/90 *Francovich v Italian State* [1993] 2 CMLR 66; Cases C-46/93 and C-48/93 *Brasserie du Pêcheur SA v Federal Republic of Germany* and *R v Secretary of State for Transport, ex parte Factortame Ltd (No. 4)* [1996] All ER (EC) 301. For discussion of the application of the principle of state liability in the context of environmental law cases, see: Hilson, "Community Rights in Environmental Law: Rhetoric or Reality?", in Holder (ed), note 124 *supra*, at pp 64-68.

¹²⁹ (1999) Cm 4444, p 20. It should be noted that the Concordat is not legally binding; it is intended to be binding in honour only.

aquatic environment also raise a number of interesting issues. On the face of it, these new powers enable a higher level of protection for Northern Ireland's aquatic wildlife and the population in general. In particular, the Department is given the power to respond to emerging scientific data concerning the potential or actual impact of new or existing pollutants on public health or aquatic wildlife. In reality, however, the Government's commitment in this regard is qualified. Paragraph 5(6) provides that where the Department takes action in relation to a discharge consent as a result of a review conducted for the purposes of protecting public health or fauna and flora, the holder of the consent will be compensated for any loss or damage sustained as a result of that action. However, the Department will not be liable to pay compensation if its action in relation to the consent is within the confines of action permitted under paragraph 5(4) *and either* the action is taken: (a) as a result of a change of circumstances that was not reasonably foreseeable when the consent was issued or most recently renewed, *or* (b) on the basis of material information¹³⁰ which was not reasonably available to the Department when the consent was issued or last renewed.

The conditions laid down in paragraph 5(6) effectively seek to ensure legal certainty, protect property rights and give expression to the principle of legitimate expectation - which in this context dictates that a recent consent will not be changed unless circumstances change markedly. Although it is unclear where the burden of proving the above conditions lies (either with the Department or the person seeking compensation, or both), they will undoubtedly operate as a 'double-edged sword'. On the one hand, they will ensure that compensation is paid only in strictly defined circumstances. However, on the other hand, to avail of its new powers of review, the Department must ensure that it acts within the confines of powers under paragraph 5(4), and also that the potential impact of discharges on public health and aquatic organisms are considered as fully as is reasonably possible when making the initial decision to grant a discharge consent. In particular, the Department must ensure that it is as up-to-date as is reasonably possible on the scientific data concerning the impact of discharges (acting alone, or in combination with other discharges or other matter) on health and wildlife that are dependant on an aquatic environment. In effect, the availability of these powers, depends ultimately on the DoE(NI)'s capacity to consider the full environmental impact of each application for a discharge consent.

(iv) The Transfer of Consents

Consistent with the principle of better regulation and in line with provisions governing discharge consents in Great Britain,¹³¹ the Water (NI) Order 1999 introduces new controls on the transfer of a consent granted by the DoE(NI). The objective of these new provisions is to ensure a seamless transfer of responsibility for the discharge. A discharge consent will be considered to be part of the holder's personal estate in the event that the holder dies and will therefore vest in his or her personal representatives. In the event of a

¹³⁰ Information is defined as being "material" in relation to a discharge consent, if "it relates to any discharge made or to be made by virtue of the consent, to the interaction of any such discharge with any other discharge or to the combined effect of any matter discharged and any other matter".

¹³¹ Schedule 10, para 11 of the WRA 1991.

bankruptcy order against the holder, the consent will be regarded as forming part of the bankrupt's estate and will vest as such in the trustee of the bankruptcy. Neither the holder of a consent, nor the holder's personal representatives or bankruptcy trustee may disclaim the consent. The holder of a discharge consent may transfer the consent to a person who intends to carry on the discharge in place of the holder.¹³² Where a consent is transferred to another holder, or vests under paragraph 8 personal representatives or trustee, the consent will take effect on and after the date of the transfer or vesting as if it had been granted to that person; the same conditions governing the discharge will apply. The Order also imposes strict rules requiring notification to the DoE(NI) of the fact of a vesting or transfer. The person from whom the consent transferred must notify the Department of the transfer within twenty-one days of the beginning of the transfer, whereas a vesting must be notified to the DoE(NI) within fifteen months of the date on which the vesting began, by the person in whom the consent vested. A failure to notify a vesting of the consent within the time period will result in the consent ceasing to have effect. A failure to comply with the notification requirements is an offence; the penalty on summary conviction is a maximum fine of £5,000 and for conviction on indictment, an unlimited fine and/or imprisonment for a maximum of two years.

(v) *The Sampling of Discharges*

Another area of significant improvement in the pollution control powers conferred under the Water Order concern the Department's powers to obtain samples of water and effluent that are admissible as evidence in legal proceedings. Although not limited to the regulation of discharge consents, these new powers will considerably enhance the Department's ability to enforce consents effectively. As was the case under section 209 of the Water Resources Act 1991, the Water Act (NI) 1972 provided that, in order to be admissible as evidence in a prosecution, samples must comply with the traditional tripartite sampling regime.¹³³ The tripartite rule required the DoE(NI) to notify the occupier of the land in question of its intention to have the sample analysed and, there and then, to divide the sample into three parts – one of which was sent for analysis, one of which was given to the occupier and one which was retained for future comparison. In addition, the Court of Appeal in *CPC (United Kingdom) Limited v National Rivers Authority*¹³⁴ ruled that samples taken by automatic continuous water monitors were inadmissible thereby requiring the regulator to base prosecutions for water pollution offences on samples that had been obtained manually.¹³⁵ Aside from the practical inconvenience involved in obtaining manual samples, taking a tripartite sample cost five times as much as a routine sample.¹³⁶ The Environment Act 1995 resolved the problem in England and Wales by repealing section 209, although this change was vigorously opposed by

¹³² Para 8.

¹³³ Section 25(5).

¹³⁴ [1995] Env LR 131

¹³⁵ In effect, the Court ruled that samples must be physically separated and isolated from the body of water and placed in a container. For a general discussion of the tripartite rule see: Hughes, note 60 *supra* at pp 576-577.

¹³⁶ Bell, *Ball & Bell on Environmental Law*, (4th ed, 1997) at p 469.

industry.¹³⁷ Section 111 of the Environment Act 1995 now simply provides that information provided or obtained pursuant to a consent condition is admissible, including information provided or obtained or recorded by means of “any apparatus” – this apparatus is also presumed to be accurate unless the contrary is shown.¹³⁸ Consequently, the Act dispensed with both the cumbersome tripartite regime and the need to ‘take’ samples manually. Although the difficulties associated with the tripartite sampling rule had been the subject of considerable criticism in England and Wales prior to the adoption of the Environment Act 1995, the 1993 *Review of the Water Act (NI) 1972* did not address the equivalent weakness in the DoE(NI)’s sampling powers under the Act. Instead the Consultation Paper focused on the problem of applying the tripartite regime to the taking of composite samples under the EC Urban Waste Water Treatment Directive 91/271.¹³⁹ Once again the delayed adoption of the Water (NI) Order 1999 produced a beneficial effect in that the Order, as finally made, does remedy this weakness in the DoE(NI)’s sampling powers. Article 27 brings Northern Ireland into line with the position under the Environment Act 1995 and includes the rebuttable presumption that the apparatus is accurate. In effect, article 27 will enable the Department to use automatic sampling machines to provide ‘legal’ samples and reduce the likelihood that polluters will evade liability as a result of evidential technicalities.¹⁴⁰

(vi) Powers of Entry and Inspection

Article 25 of the Water Order considerably increases the DoE(NI)’s powers to enter and inspect premises for the purpose of exercising its pollution control powers under the Order and ensuring that the terms of the Order are being complied with. Unfortunately, however, this is one area where the changes wrought do not give the Department the very wide powers available to the Environment Agency under section 108 of the Environment Act 1995. One of the most important improvements in the Department’s powers in this respect is the extension of the term premises to include “mobile plant”, which includes fish cages.¹⁴¹ Although aquaculture in Northern Ireland is currently under-developed when compared to the Republic of Ireland, this is

¹³⁷ *Ibid*, at p 470.

¹³⁸ Section 111(3).

¹³⁹ OJ L135, 30 May 1991. Composite samples involve sampling over a 24-hour period, bulking the 24 parts and analysing the composite sample – hence sampling by an individual officer would be time-consuming and expensive. See also the *Review of the Water Act (NI) 1972: A Consultation Paper* (December 1993, DoE(NI) and DANI) at pp 13, para 7.1.

¹⁴⁰ The definition of “apparatus” contained in Article 3(1) of the Water Order is identical to the definition contained in section 111(5) of the EA 1995. A number of automatic sampling devices have been developed for continuous monitoring. For a description of ‘Merlin’, ‘Sherlock’ and ‘Cyclops’ see: Mumma, “Use of Compliance Monitoring Data in Water Prosecutions” (1993) 3 *Journal of Environmental Law* 19.

¹⁴¹ Article 25. The Order does not provide a direct definition of the term “premises”; however article 25(5) includes mobile plant in the definition of premises that may be entered and the explanatory document accompanying the Order highlights mobile plant as being included in the definition. Mobile plant is defined in article 3(1) as meaning plant which is designed to move or be moved whether by road or otherwise.

an industry that is likely to expand considerably in the future.¹⁴² Another important change is the extension of the Department's powers of inspection from simply "inspecting and surveying" under the 1972 Act¹⁴³ to include measuring, testing and taking away samples of water, effluent, land or articles.¹⁴⁴ In doing so, the Order brings the DoE(NI)'s inspection powers broadly into line with the Environment Agency's powers, but there are also regrettable gaps in the Department's powers in this regard. In particular, the DoE(NI), unlike the Environment Agency, is not explicitly conferred with the power to take photographs, nor is it given the power to take samples of air and substances. Although the Order gives the DoE(NI)'s new powers to install to and keep monitoring or other apparatus on a premises, to fix sampling points and to carry out experimental borings or other works on the premises, this reform is somewhat half-hearted. Unlike the Environment Agency, the Department is not given power to direct that any premises it has entered or anything in them must be left undisturbed for so long as is necessary to conduct the examination or investigation. It goes without saying that this omission may limit to a considerable degree the Department's ability to use its new powers. Similarly, the DoE(NI), unlike the Agency, is not given power to take, dismantle, damage or destroy an article or substance found on a premises which is likely to cause pollution to the environment or harm to human health. Furthermore, the Department is not given the important powers granted to the Agency to require the provision of facilities and assistance with respect to matters or things within another person's control that is necessary to exercise the powers of entry and inspection. Nor is the Department conferred with the Agency's power to require any person who may be able to provide information relevant to the investigation to answer questions and to sign a declaration testifying to the truth of their answers. Last, but not least, the Water Order fails to bring the DoE(NI)'s emergency powers into line with those conferred on the Environment Agency. Whereas the Agency may make a forced entry at any time in cases of emergency, the Department must obtain a warrant to enter, and strangely, is only empowered to enter at a reasonable hour.

SECTION II

AN AGENDA FOR FURTHER REFORM

The analysis thus far of the Water (NI) Order 1999 makes clear that this new legislation represents an important, albeit overdue, milestone in the modernisation of water pollution law and policy in Northern Ireland. When it comes into force, the core legislative framework governing water pollution control will implement principles and policies that are now considered to be fundamental to contemporary national and EC environmental law. However, although Northern Ireland is on the verge of gaining the means by which to

¹⁴² See generally, *Northern Ireland Aquaculture Development*, October 1995: Study Commissioned by Northern Ireland Seafood. See also the following publications produced by the Irish Sea Fisheries Board, *Seafood Industry Agenda 2000-2006: Realising the Regional Potential of the Irish Seafood Industry* (March 1999) and *Aquaculture and the Environment*.

¹⁴³ Section 23(1).

¹⁴⁴ Article 25(2)(b).

afford a high level of legal protection to its aquatic environment, the process of modernisation is far from complete.

Despite being totally at odds with national and EC policy on pollution control, the Water (NI) Order 1999 makes no attempt to resolve the poacher-gamekeeper conflict that continues to undermine the DoE(NI)'s credibility as regulator in the field of water pollution. In addition, significant disparities remain in the level of protection afforded to the freshwater environment in Northern Ireland and that extended to its counterparts in England and Wales. Although parity with England and Wales may become less important as a result of devolution in the field of environmental protection, a decision to tolerate weaker standards in this jurisdiction may guarantee Northern Ireland's future as a pollution haven within the United Kingdom. To compound these problems the Northern Ireland Comptroller and Auditor General has recently reported a series of fundamental shortcomings in the DoE(NI)'s administration and enforcement of the existing statutory framework governing the aquatic environment in Northern Ireland.¹⁴⁵ The Auditor General's highly critical report documents a scene of almost endemic regulatory failure thus raising serious concerns as to the Department's ability to engage with the more rigorous controls introduced by the new Water Order. Last, but not least, there is a failure to grapple with the forthcoming tidal wave of reform in the field of EC water law. The European Union is currently preparing to enact a new Water Framework Directive,¹⁴⁶ which represents a thorough restructuring of EU policy on the aquatic environment and will set objectives for water protection well into the next century. Although the adoption of the Water (NI) Order 1999 provided an ideal opportunity to lay the foundations for implementing the terms of this Directive in Northern Ireland, little effort was made to anticipate the demands of this new phase of reform.

In December 1999 legislative and executive powers were devolved to the Northern Ireland Assembly under the Northern Ireland Act 1998.¹⁴⁷ With only two exceptions,¹⁴⁸ all responsibility for environmental protection was devolved to the Assembly, including power to make law and policy on the issue of water pollution subject of course to the proviso that it must act in accordance with international and EC law.¹⁴⁹ Although at the time of writing the legislative and executive powers of the Assembly have been suspended

¹⁴⁵ *Control of River Pollution in Northern Ireland*, Report by the Comptroller and Auditor General for Northern Ireland; HC 693, Session 1997-98.

¹⁴⁶ Council Text, *The European Union Council, Common Position*, 19 February, 1999; Commission Texts, COM(1999) 271 final, COM(97) 49 final, COM(98) 76 final; European Parliament Texts, *Parliament's First Reading Amendments A4-0261/98*, COM(97) 49 final, COM(98) 76 final.

¹⁴⁷ c.47.

¹⁴⁸ Schedule 3, para 39 provides that the subject matter of section 3(5) to (7) of the Environmental Protection Act 1990 (emission limits and quality objectives) and the environmental protection technology scheme for research and development in the United Kingdom are reserved matters. Under section 3 of the EPA 1990, the Secretary of State has the power to make national plans establishing emission limits for the release of any substance into the environment or for the progressive improvement in quality objectives and standards.

¹⁴⁹ Section 6 of the Northern Ireland Act 1998.

and the North/South and British-Irish elements have been stalled,¹⁵⁰ if they are revived the aquatic environment would be the subject of governance through a complex web of structures. First and foremost, responsibility for the aquatic environment will lie with four of the new departments of the Northern Ireland Assembly. Pollution control, and therefore the administration and enforcement of much of the Water Order, will be dealt with by the Department of the Environment. Responsibility for the provision of water falls within the remit of the Department for Regional Development. The development of inland waterways and fisheries will be dealt with by the Department of Culture, Arts and Leisure, while the Department of Agriculture and Rural Development will deal generally with the issue of farming and environmental policy, which will inevitably impact on the freshwater environment. However, it is also worth noting that the aquatic environment could also come within the remit of the Office of the First Minister and Deputy First Minister, for example, as an issue of policy innovation. Each of the new departments will be assisted and advised by a 'statutory committee' with power to play a scrutiny, policy development and consultation role in relation to its department and a role in the initiation and opposition of legislation.¹⁵¹ Secondly, the management of the aquatic environment will be the subject of all three tiers of the all-Ireland dimension to Northern Ireland devolution; namely the North/South Ministerial Council,¹⁵² the cross-border implementation bodies¹⁵³ and the (initial) six areas identified for co-operation between existing government structures, North and South.¹⁵⁴ Thirdly, the environment comes within the remit of the British-Irish Council established to promote the harmonious and mutually beneficial development of relationships among the peoples of the United Kingdom and Ireland.¹⁵⁵ The British-Irish Council will exchange

¹⁵⁰ Devolution was suspended with effect from 12 February 2000 by the Northern Ireland Act 2000, c. 1.

¹⁵¹ The system of statutory committees (as envisaged by Strand 1, para 9 of the Belfast Agreement) was established under section 29 of the Northern Ireland Act 1998.

¹⁵² Strand 2, paragraph 1 of the Belfast Agreement refers to the establishment of a North South Ministerial Council designed to bring together Ministers from Northern Ireland and the Irish Government to develop consultation, co-operation and action on an all-island basis on matters of mutual interest. See also section 52 of the Northern Ireland Act 1998. It should be noted that the Government's Memorandum of Understanding and 4 Concordats of October 1999 (Cm 4444) deal, *inter alia*, with North/ South Council matters.

¹⁵³ Six cross-border implementation bodies were established by international agreement between the British and Irish Governments for the purpose of implementing the policies agreed by the Ministers in the North/South Ministerial Council. See: the (Irish) British Irish Agreement Act and the North South Co-operation (Implementation Bodies) (NI) Order 1999 (SI 1999/859). The six bodies came into existence on the 2 December, 1999. Two of the six bodies have a specific role to play in relation to the freshwater environment; namely, Waterways Ireland and The Foyle, Carlingford and Irish Lights Commission.

¹⁵⁴ These six (initial) areas of co-operation are in addition to the six Implementation Bodies. The 'environment' is identified as one such area of co-operation, specifically for present purposes, research into environmental protection and water quality management.

¹⁵⁵ See sections 52 and 53 of the Northern Act 1998 and Strand 3 of the Belfast Agreement. It should be noted that the Government's Memorandum of

information, discuss, consult and use best endeavours to reach agreement on matters of mutual interest.¹⁵⁶ Environmental matters have been identified as one of the priority areas to be addressed by the British-Irish Council.¹⁵⁷ Because primary responsibility for parliamentary scrutiny in devolved matters passes to the Assembly under the Northern Ireland Act 1998,¹⁵⁸ it is unlikely that independent scrutiny of the Assembly in the field of water pollution control will come in the form of another House of Commons Select Committee Report such as the seminal ‘Rossi Report’ in 1990 which was largely comprised of MPs external to Northern Ireland.¹⁵⁹ Similarly, with EC policy on water pollution shifting from its traditionally rigid approach to control towards the conferral of greater discretion to Member States in the implementation of water Directives,¹⁶⁰ the Assembly and its departments will have more freedom in tailoring the transposition of Directives to the specific conditions pertaining in Northern Ireland. Consequently, if devolution in the form outlined above is revived, much more responsibility will be given to the local political framework to have the will and the confidence to recognise the shortcomings still latent in the regulatory system governing water pollution, the weaknesses of self-regulation and the need for independent scrutiny.

The purpose of Section II of this article is to set out an agenda for further reform of the regulatory framework governing pollution of the freshwater environment in Northern Ireland. Although at the time of writing devolution has been suspended, regrettably the same cannot be said of the need for further reform in this field. While it is possible that the devolution arrangements laid down in the Northern Ireland Act 1998 may remain suspended for an indefinite period, the process of economic regeneration will continue as long as the ceasefires hold. However, to have any hope of ensuring that economic regeneration occurs in concert with – not at the expense of – the effective protection of the freshwater environment, it is vital that immediate steps are taken to address the remaining weaknesses underlying the system of regulation governing water pollution. Indeed, even if the political situation destabilises and political violence re-emerges, the process of reform in the field of water pollution control must not be suspended. As the House of Commons Select Committee on the Environment stated in their 1990 report on *Environmental Issues in Northern*

Understanding and 4 Concordats of October 1999, note 129 *supra*, deal, *inter alia*, with British Irish Council matters. The British-Irish Council will be comprised of representatives of the British and Irish Governments, of the devolved institutions of Northern Ireland, Scotland and Wales and of the Isle of Man and the Channel Islands.

¹⁵⁶ *Devolving Power to the People of Northern Ireland*, Executive Information Service of the Office of the First Minister and Deputy First Minister.

¹⁵⁷ *Ibid.*

¹⁵⁸ Northern Ireland Affairs, *First Special Report*; HC 178, Session 2000.

¹⁵⁹ *Environmental Issues in Northern Ireland*, HC 39, Session 1990-91. The Report became known as the ‘Rossi Report’ after its chairman, Sir Hugh Rossi. The Report had a fundamental impact on the modernisation of environmental law in Northern Ireland; its findings and their effect in the field of water pollution laws are addressed in Part I of this article, (2000) 51 *NILQ* 65 *et seq.*

¹⁶⁰ Howarth, “Accommodation without Resolution? Emission Control and Environmental Quality Objectives in the Proposed EC Framework Directive” (1990) 1 *Environmental Law Review* 6.

Ireland:¹⁶¹ “the existence of the ‘troubles’ does not remove the need for dealing effectively with the environmental problems of the Province which are as great as, and in some respects greater than, those of the rest of the United Kingdom.”¹⁶² If the future of the aquatic environment is to be adequately safeguarded, it is imperative that the Government’s commitment to further reform is not contingent upon progress being made in reaching a sustained constitutional settlement. The specific nature and scope of the remaining regulatory problems will now be discussed in more detail.

Resolution of the Poacher-Gamekeeper Dilemma

The most important shortcoming of the Water (NI) Order 1999 is undoubtedly its failure to resolve the poacher-gamekeeper conflict that has long undermined the DoE(NI)’s role as regulator in the field of water pollution control. As was explained in Part One¹⁶³ of this analysis, the DoE(NI) is responsible not only for administering and enforcing the system of regulation governing freshwater pollution,¹⁶⁴ but also for treatment and supply of drinking water and the collection, treatment and disposal of sewage and waste water.¹⁶⁵ Because the Department’s water and sewage treatment works are a source of water pollution,¹⁶⁶ the DoE(NI) operates as both the regulator and a polluter of water. Despite the fact that national and European policy on pollution control since 1990 has necessitated the separation of regulatory and operational functions,¹⁶⁷ and despite the recommendation of the House of Commons Environment Select Committee in 1990 that this conflict of interest be resolved,¹⁶⁸ the *Review of the Water Act (NI) 1972*¹⁶⁹ published in 1993 did not address this fundamental issue. Nevertheless, several of the environmental interests groups responding to the consultation paper strongly urged the Government to address this issue as a matter of urgency.¹⁷⁰ Similarly, members of the Standing Committee on Delegated Legislation expressed considerable concern that the Government had not

¹⁶¹ Note 159 *supra*.

¹⁶² Note 159 *supra*, at p vii, para 1.

¹⁶³ (2000) 51 *NILQ*, 65 *et seq.*

¹⁶⁴ Under the Water Act (NI) 1972 – this power is exercised by the Environment and Heritage Service (a Next Steps Agency to the Department).

¹⁶⁵ Under the Water and Sewerage Services (NI) Order 1973 (SI 1973/70) – this function is exercised by the Water Service (a Next Steps Agency to the Department).

¹⁶⁶ Note 145 *supra*, at Part 4.

¹⁶⁷ *This Common Inheritance: Britain’s Environmental Strategy*, (1990) Cm 1200, at p 139 para 10.14. See also the discussion of this issue in Part One of this article, (2000) 51 *NILQ* at 65 *et seq.*

¹⁶⁸ Note 159 *supra*, at p xxiv, para 78.

¹⁶⁹ December 1993 (DoE(NI) and the Department of Agriculture (NI)).

¹⁷⁰ Although not all responses to the consultation exercise were made available, the following interest groups strongly urged the Government to separate the DoE(NI)’s regulatory and operational functions: Northern Ireland Environmental Link (on behalf of International Tree Foundation, Lenadoon Environmental Forum, Mountaineering Council of Ireland, Mournes Heritage Trust, Ulster Archaeological Society, Ulster Architectural Heritage Society, Ulster Foundation for the Preservation of the Countryside, Upper Faughan River Trust, and the Wildfowl and Wetlands Trust); Friends of the Earth and The Royal Society for the Protection of Birds.

used the making of the Order as an opportunity to resolve this fundamental problem.¹⁷¹ Whereas in England and Wales the structural and substantive dimensions of the regulatory system were reformed in one fell swoop by the Water Act 1989, in Northern Ireland the Government decided instead to deal with these issues separately and slowly. Having waited a decade to bring the legislation governing water pollution in Northern Ireland broadly into line with that operating in Great Britain, it may be several more years before the structures responsible for delivering water pollution control are finally reformed. In the meantime, the report published by the Northern Ireland Auditor's Office in 1998 concerning the control of river pollution in Northern Ireland raises considerable concerns about the lax standards of control currently applied to discharges from DoE(NI) sewage treatment works.¹⁷² The report pointed out that the DoE(NI)'s Crown Immunity not only prevented the Environment and Heritage Service (EHS) prosecuting the Water Service for pollution caused by discharges from its sewage treatment works, but also that the standards set by the EHS for such discharges are significantly less stringent than those currently applied by the Environment Agency in England and Wales. Although discharge standards for sewage treatment works in England and Wales were relaxed for a period of time prior to privatisation to enable capital works to be carried out, by 1994 the Government reported a "major, sharp improvement since 1986 in the compliance by sewage works with standards required".¹⁷³ It also reported similar "steady improvement" taking place in Scotland.¹⁷⁴ In contrast, the Auditor General reported that EHS standards for sewage treatment works in Northern Ireland are currently even more relaxed than the relaxed standards imposed on the Regional Water Authorities in England and Wales prior to privatisation.¹⁷⁵ Furthermore, it was pointed out that although the data for 1996 indicated that 76% of treatment works would comply with the standards set by the EHS, the use of these standards for the purposes of public accountability "may be misleading, creating an illusion of compliance with high quality standards, which is not, in fact the case."¹⁷⁶ In response, the EHS stated that in the absence of the regulatory powers made available in the wake of the separation of regulatory and operational functions in Great Britain, the present regime was as close as possible to the system operating in England and Wales.¹⁷⁷

Despite the fundamentally weak nature of the control imposed on pollution from DoE(NI) sewage and water treatment works, the process of addressing the Department's conflicting regulatory and operational roles was only recently launched. In November 1998 the DoE(NI) published *Water and Sewerage Services in Northern Ireland: A Consultation Paper*, which sets out the Government's proposals for restructuring the provision of water and

¹⁷¹ In particular, Mr Malcolm Moss and Mr Clifford Forsythe raised this issue at length; Third Standing Committee on Delegated Legislation, House of Commons, 2 March 1999.

¹⁷² Note 145 *supra* at pp 47-55.

¹⁷³ *Sustainable Development: The United Kingdom Strategy*, (1994) Cm 2426. Chapter 8, at p 62, para 8.38.

¹⁷⁴ *Ibid.*

¹⁷⁵ Note 145 *supra* at pp 49-50.

¹⁷⁶ *Ibid.*, p 50 at para 4.24.

¹⁷⁷ *Ibid.*

sewerage services and the regulatory structure governing the management of water resources in Northern Ireland. Although the DoE(NI) recommended splitting its policing and operational functions,¹⁷⁸ it is interesting to note that the Consultation Paper characterised this weakness in the regulatory system as perceived rather than real, and as one which necessitated reform in order to increase public confidence rather than ensure compliance with long established national policy on the regulation of water pollution.¹⁷⁹ In August 1999, the Minister for the Environment (Lord Dubbs) issued a press statement concerning the response to the consultation process. No clear consensus emerged from the process concerning the regulatory structures; however, considerable support was expressed for either an Ombudsman or an independent regulator.¹⁸⁰ Lord Dubbs reported that officials had been instructed to develop further proposals for reform based on the outcome of the consultation process; however, it was anticipated that the final decision concerning the future of the water industry and regulatory structures would be taken by the Northern Ireland Assembly. Although the Assembly did not address this issue directly during its brief existence, one element of the poacher-gamekeeper problem was temporarily resolved as a result of the allocation of responsibilities amongst the newly established Northern Ireland departments. Whereas responsibility for pollution control remained with the newly created Department of the Environment, responsibility for the functions formerly carried out by the Water Service were re-allocated to the new Department for Regional Development.¹⁸¹ Consequently, if the Assembly's devolved powers were restored, the new departmental structures would effectively separate the regulatory and operational functions relating to water. However, the true benefit of this change may be more apparent than real. As a result of its Crown Immunity, the Department for Regional Development could not be brought within the controls of the discharge consent system nor could it be held accountable for any water pollution caused by discharges from its water and sewage treatment works. If Northern Ireland is to have a truly transparent system of water pollution control in which the public can be confident, it is essential that the environmental regulator deciding when and how to enforce is truly independent and has no responsibility for operational functions. Although the DoE would no longer be responsible for operational matters if devolution were revived, as long as the Environment and Heritage Service remains as a Next Steps agency within the Department, it cannot be said to be a truly independent environmental regulator.¹⁸² In addition, as long as those responsible for water and sewerage services enjoy the protection of Crown Immunity, even the possibility of private prosecutions is ruled out. Consequently, if the aquatic environment in Northern Ireland is to be afforded the level of protection associated with modern standards of

¹⁷⁸ P 14.

¹⁷⁹ *Ibid*, at para 4.7.

¹⁸⁰ See also the discussion of the consultation process in the *Belfast (Business) Telegraph*, Tuesday, August 17, 1999, pp 10.

¹⁸¹ Note 156 *supra*.

¹⁸² The nature of the EHS Next Steps arrangements are discussed in K. Morrow & S. Turner, "The More Things Change, the More they Stay the Same? Environmental Law, Policy and Funding in Northern Ireland" (1998) *Journal of Environmental Law* pp 44-46.

environmental regulation operating elsewhere in the United Kingdom, further reform is essential.

Removal of Continuing Legislative Disparity

The second major shortcoming of the Water Order is its failure to bring the core legislative framework governing pollution control in Northern Ireland into full parity with that applying in England and Wales. Although a restoration of the Assembly's devolved powers would enable different routes to be taken to ensuring environmental protection, it is submitted that the disparities in question are unjustified and will expose freshwater in Northern Ireland to a greater risk of pollution than elsewhere in the United Kingdom. Each of these shortcomings will be addressed in turn.

(i) The General Obligations Imposed on the DoE(NI)

Article 4(1), which sets down the DoE(NI)'s general duties under the Water (NI) Order 1999, is without doubt one of the most revealing areas of disparity between the legislation governing water pollution in Northern Ireland and that applying in England and Wales. Article 4(1) provides that the Department shall "promote" both the conservation of Northern Ireland's water resources and the cleanliness of its inland, coastal and ground waters. Article 4(2) goes on to provide that in carrying out its duties in each context, the Department must "have regard to" a series of interests. They are: the needs of industry and agriculture; the protection of fisheries; the protection of public health; the preservation of amenity and the conservation of flora and fauna; and the conservation of geological or physiographical features of special interest and any feature of archaeological, historical, architectural or traditional interest. Except for the addition of the latter interest, the DoE(NI)'s general duties in relation to pollution control or water resource conservation have remained virtually unchanged despite the far-reaching changes that have occurred in field of water law and policy during the twenty-seven years since the enactment of the Water Act (NI) 1972.

In stark contrast, the statutory aims and duties imposed on the Environment Agency in England and Wales not only make a direct link with the Government's commitment to achieving sustainable development, but also reflect the contemporary policy emphasis on pollution prevention and conservation of nature and the countryside. Section 4 of the Environment Act 1995 states that the "principal aim" of the Environment Agency is to protect or enhance the environment so as to make a contribution towards attaining the objective of sustainable development. Although the Environment and Heritage Service *Corporate Plan 1998/2001* states that the Service "aims to be seen as the main protector of the environment in Northern Ireland and a significant influence and lead player in the promotion of sustainable development", the Water Order fails to forge a statutory link between the DOE(NI)'s pollution control function and its role in promoting sustainability.¹⁸³ Section 5 of the Environment Act 1995 gives explicit

¹⁸³ Section 33(1) and (2) of the Environment Act 1995 empower the Secretary of State to give guidance to the Scottish Environment Protection Agency concerning the aims and objectives which he considers appropriate for SEPA to pursue in the

expression to the principle of pollution prevention that is now fundamental to EC and national legislation on pollution control. Under section 5, the Environment Agency is required to exercise its pollution control powers for the purpose of preventing or minimising, or remedying or mitigating the effects of, pollution of the environment.¹⁸⁴ Although one of the major innovations of the Water (NI) Order 1999 is the conferral of stronger powers to prevent water pollution, the general duties imposed on the DoE(NI) concerning the exercise of its pollution control powers do not even allude to this considerably expanded dimension of the Department's regulatory function.¹⁸⁵ By contrast, it is interesting to note that article 40(1) explicitly requires the Department of Agriculture (NI) "to have regard to the prevention of pollution" in the exercise of its functions under the Order concerning inland water navigation and recreation.¹⁸⁶

Section 6 of the Environment Act 1995 sets out the Environment Agency's general obligations with respect to water. In this regard the Agency is required to promote "to the extent that it considers desirable": (a) the conservation and enhancement of the natural beauty and amenity of inland and coastal waters and the land associated with such waters and (b) the conservation of flora and fauna which are dependant on an aquatic environment.¹⁸⁷ The Environment Agency is also placed under a general duty to take action to secure "the proper use of water resources".¹⁸⁸ Although the Northern Ireland Order requires the DoE(NI) to promote the conservation of water resources, the Department is not placed under a specific duty to secure the proper use of water resources in Northern Ireland.¹⁸⁹ And while the nature conservation duties contained in section 6 are not particularly onerous in that action is only required when the Agency deems it to be desirable, the commitment to conservation is clearly much stronger than that expressed in article 4 of the Water (NI) Order 1999. Whereas the Environment Agency is required to "promote" the conservation and enhancement of the natural beauty and amenity of water and "promote" the conservation of flora and fauna that are dependant on an aquatic environment, the DoE(NI) is only required to "have regard" to the

performance of its functions. However, any such guidance must include guidance concerning SEPA's contribution towards attaining sustainable development.

¹⁸⁴ Similarly, the Scottish Environment Protection Agency is required under section 33(1) of the Environment Act 1995 to exercise its pollution control powers "for the purpose of preventing or minimising, or remedying or mitigating the effects of pollution on the environment."

¹⁸⁵ It is interesting to note that the DoE(NI) is obliged to have regard to the prevention of pollution when exercising its powers under article 23 to (i) carry out engineering or building operations under Part II of the Order; (ii) vary the flow of water or (iii) discharge water into any waterway.

¹⁸⁶ It should also be noted that DANI's powers under the Water Order concerning water recreation will be reallocated to the Department of Culture, Arts and Leisure.

¹⁸⁷ Section 6(1). The Scottish Environment Protection Agency is placed under similar obligations under sections 32 and 34 of the Environment Act 1995.

¹⁸⁸ Section 6(2).

¹⁸⁹ The Department's powers to manage the water resource have however been strengthened under the Water Order through the conferral of new powers to control the abstraction of water by means of a licensing system and the imposition of abstraction charges under articles 20 to 22.

preservation of amenity and the conservation of flora and fauna. The Department is not required even “to have regard” to the “enhancement” of the natural beauty and amenity of such waters. Nor is the DoE(NI) specifically required to have regard to the conservation of land associated with inland and coastal waters, although it is placed under a duty to have regard to the conservation of geological or physiographical features of special interest and any feature of archaeological, historical, architectural or traditional interest. It is interesting, however, to note that under article 40(1), DANI is required to “have regard to the conservation *and enhancement* of flora and fauna” in exercising its functions under the Order concerning inland water navigation and water recreation. Furthermore, the Environment Agency’s obligation to promote the conservation of the aquatic environment and its flora and fauna is also not placed in direct competition with the needs of other conflicting interests such as industry and agriculture which, under the Water (NI) Order 1999, arguably occupy a higher position in the hierarchy of interests to which the DoE(NI) must have regard. The greater emphasis on nature conservation reflected in the Environment Agency’s general duties is further reinforced by section 7 of the Environment Act 1995. Under section 7 the Environment Agency is required to exercise its non-pollution powers (that is, flood defence and fisheries) so as to “*further*” the conservation and enhancement of natural beauty, flora and fauna, and geological or physiographical features of special interest. When exercising its pollution control functions, the Agency is required to have regard to the “*desirability*” of conserving and enhancing such features.

During the consultation process surrounding the review of the Water Act (NI) 1972 the Ulster Wildlife Trust¹⁹⁰ and the Royal Society for the Protection of Birds (NI)¹⁹¹ both emphasised the need to strengthen the Department’s general commitment to environmental protection and specifically to nature conservation. The UWT pointed out that the list of interests to which the Department was to have regard implied an order of priority, and combined with the term “shall have regard to” placed flora and fauna in a very weak position. Both the UWT and the RSPB recommended that the DoE(NI)’s general nature conservation duties under the proposed Water Order should be at least as strong as those operating in England and Wales. In particular, the RSPB pointed out that the Secretary of State for the Environment, Transport and the Regions, the Secretary of State for Wales, the Director General of Water Services, the Environment Agency and British Waterways all have a duty “to further the conservation and enhancement of natural beauty and the conservation of flora, fauna and geological or physiographical features of interest.” The RSPB also pointed out that the need to further nature conservation is underlined by the United Kingdom’s biodiversity commitments. The biodiversity process introduces a much more definite role for public bodies in contributing to specific environmental targets, including the enhancement of species and habitats. Hence, the RSPB

¹⁹⁰ The UWT made one response during the consultation process on the 24 March 1994 following the publication of the Consultation Paper in December 1993.

¹⁹¹ The RSPB(NI) made two responses during the consultation process. The first was made in response to the publication of the Consultation Paper in December 1993; the second was made in August 1998 in response to the publication of the proposal for a draft Order.

argued that the requirements of the biodiversity process to meet targets through enhancement work should, therefore, be reflected by legislation where appropriate. Although the RSPB welcomed the wording contained in article 40(1) concerning DANI's general obligations under the Order, they regarded the continued use of the phrase "have regard to" as undesirable in that it suggested that a choice could be made to conserve, enhance or not, which in turn could lead to key opportunities being lost or habitats and wildlife being lost. The RSPB stated that all opportunities for protecting or enhancing nature conservation should be taken if national and local biodiversity targets are to be met; hence they recommended that article 40(1) should be re-worded to reflect this position more clearly. In addition, the RSPB expressed their concern that the DoE(NI) was not placed under a general duty to secure the proper use of water resources. The Society pointed out that such an overarching duty would require the Department to assess the need for new water resource development and ensure that the most appropriate schemes are licensed, while taking into account the impact on the environment and other users.

The review of the Water Act (NI) 1972 presented the Government with an important opportunity to show the extent of its commitment to embracing the recommendation made by the House of Commons Environment Select Committee that, in the absence of an Environment Agency for Northern Ireland, environmental protection would be given a higher profile in the work of the DoE(NI).¹⁹² Despite the compelling nature of the arguments raised by the UWF and RSPB and the specific recommendation by the House of Commons Committee that the environment in Northern Ireland be afforded equal legal protection to that in the rest of the United Kingdom, the Government remained unmoved by these recommendations. Although the principal thrust of the Water (NI) Order 1999 is to considerably strengthen the pollution prevention powers available to the DoE(NI), the general duties imposed on the Department are inconsistent not only with the general duties imposed on the regulator in England and Wales, but also with the largely progressive nature of the Order itself. In addition, the Department's nature conservation duties are inconsistent not only with the United Kingdom's biodiversity commitments, but also with the proposed Biodiversity Strategy published in June 1999 by the Government's advisory group (the Northern Ireland Biodiversity Group) to guide the delivery of the United Kingdom's commitments in this jurisdiction.¹⁹³ While it can certainly be argued that the imposition of strong general obligations is no guarantee that the body will, in practice, show a strong regard for the environment, it is submitted that, in the absence of an independent environmental regulator in Northern Ireland, the nature of the general duties imposed on the DoE are at least of considerable symbolic importance.

¹⁹² *Environmental Issues in Northern Ireland*; HC 39, Session 1990-91, recommendation 25, para 124. In this regard, see also, Milton, *Our Countryside Our Concern: The policy and practice of conservation in Northern Ireland*, Northern Ireland Environment Link (1990), Chapter 2.

¹⁹³ *Northern Ireland Biodiversity Strategy Proposals*, DoE(NI) June 1999.

(ii) The Obligation to Achieve Water Quality Objectives

The second significant instance of legislative disparity concerns the Government's failure to place the DoE(NI) under a statutory duty to exercise its powers under the Water Order so as to achieve statutory water quality objectives at all times in so far as it is practicable to do so. This duty is imposed on the Environment Agency under section 84 of the Water Resources Act 1991 and applies to quality objectives as they are formally established by regulations. While the Agency is not deemed to be in breach of its obligations under section 84 simply by failing to achieve the objectives, it is required to ensure that its powers to grant and review discharge consents and its preventative, remedial and enforcement powers are exercised so as to ensure that the objectives are achieved if practicable. Failure to do so would provide the basis for judicial review proceedings to enforce the Agency's duty in this regard thereby providing third parties an important means of engaging with the enforcement of this vital arm of the regulatory system.¹⁹⁴ Apart from the issue of disparity with national legislation, the Government's decision to confer the DoE(NI) with entirely discretionary powers to achieve water quality objectives is even more regrettable in light of the forthcoming enactment of the proposed EU Framework Directive on Water Quality and the highly critical findings of both the Halcrow Report in 1988 and the Comptroller and Auditor General in 1998 concerning the Department's strategy for monitoring water quality in Northern Ireland. The proposed Framework Directive specifically identifies the need for "more intensive monitoring of the extent and nature of pollution and immediate review of all relevant authorisations and discharge permits followed by action based upon the level of risk involved."¹⁹⁵ Although article 6 of the Water Order goes some way towards implementing the requirements of the Directive, the Order does not place the DoE(NI) under an obligation to conduct any follow-up action in the event of a failure to meet quality objectives by a specified date.¹⁹⁶ It is anticipated that the Framework Directive will be enacted during the year 2000.

However, of even greater immediate concern is the DoE(NI)'s failure to conduct an adequate programme of water quality monitoring. In 1988 the Halcrow Report stated that the systems employed by the DoE(NI)'s Water Quality Unit for measuring water quality were "wholly insufficient to measure the performance of the regulatory authority or to give an indication of overall water quality in the Province".¹⁹⁷ The Halcrow Report stated that

¹⁹⁴ Bell, note 136 *supra*, at, p 466. An additional issue in this regard concerns the Government's failure to impose a timetable for the implementation of water quality objectives – thus far none has been imposed on the Environment Agency. For further discussion of the environmental issues which should be addressed via water quality objectives in Northern Ireland, see the views expressed by the RSPB(NI) during the consultation process, 24 March 1994, para 5.3.

¹⁹⁵ Article 13 of the Commission's proposal, COM(1999) 271 final.

¹⁹⁶ In this regard, see also the comments submitted to the DoE(NI) by the RSPB(NI) in response to the consultation process, 13 August 1998, para 2.1.

¹⁹⁷ Paras. 1.4–1.9. The Halcrow Report was a report commissioned by the DoE(NI) and is therefore an internal document. Although the author requested a copy of the report during the writing of this paper, the Environment and Heritage Service made a decision (3 months later) to refuse access on the grounds that it is a

the length of rivers monitored for water quality was “probably less than adequate to support even the minimal length of rivers classified”.¹⁹⁸ It added, “that if Northern Ireland was to demonstrate a commitment to aquatic environment quality at least comparable to elsewhere in the United Kingdom, a much greater length of rivers should be classified.”¹⁹⁹ A decade later the Northern Ireland Auditor General confirmed the findings of the Halcrow Report. The Auditor General pointed out that the DoE(NI) failed to make any response to the Halcrow Report and did not extend the length of rivers monitored to meet the Report’s medium and long-term objectives.²⁰⁰ Furthermore, the Auditor General reported that the Environment and Heritage Service (EHS) did not collect information with sufficient frequency to properly assess progress against its stated water quality objectives.²⁰¹ The Auditor General also reported that the EHS was still monitoring a substantially lower length of river than comparable regions in the United Kingdom.²⁰² However, most damning of all was the Auditor General’s decision to record the Department of Agriculture’s contention that the monitoring programme was unrepresentative because it did not monitor smaller rural streams in which the water quality was markedly worse than in the major rivers included in the monitoring network.²⁰³ Although the EHS disputed both DANI’s submission and the findings of the Halcrow Report, the Auditor General concluded that because the EHS had not carried out any formal assessment of river monitoring needs to confirm the adequacy of its current monitoring strategy, neither criticism could be dismissed.²⁰⁴ Although the EHS has announced its decision to undertake a review of its water quality monitoring strategy which will address the concerns raised by the Halcrow Report, DANI and the Auditor General, given the DoE(NI)’s track record in relation to water quality control and the forthcoming obligations under the Framework Directive, it is regrettable that the Government did not see fit to place the Department (like the Environment Agency) under an obligation to achieve water quality objectives in Northern Ireland in so far as is practicable.

(iii) The Liability of Directors and Senior Company Officers

A third area of significant disparity between the Water Resources Act 1991 and the Water Order is the lack of provision under the Northern Ireland legislation for the imposition of criminal liability on individual directors, senior company officers and third parties for regulatory offences. Section 217 of the WRA 1991 enables such persons to be held personally liable in criminal law for *any* offence under the Act that is committed by their company with their consent or connivance, or which can be attributed to

confidential, internal report within the meaning of Regulation 5(2)(c) of the Environmental Information Regulations (NI) 1993. However, reference to Halcrow’s core findings can be found in the Report of the Comptroller and Auditor General for Northern Ireland, note 145 *supra*, at p.39.

¹⁹⁸ Report of the Comptroller and Auditor General for Northern Ireland, *ibid*.

¹⁹⁹ *Ibid*.

²⁰⁰ *Ibid*, at p 39, para 2.22.

²⁰¹ *Ibid*, at p 38, para 2..19.

²⁰² *Ibid*, at p 39, para 2.21.

²⁰³ *Ibid*, at p 40, para 2.24-2.26.

²⁰⁴ *Ibid*, at p 40, para 2.26 and p 41, para 2.27.

neglect on their part. Individual liability for corporate offences is one of the most important contemporary strategies employed in the regulation of water pollution in England and Wales and undoubtedly provides a powerful incentive to senior company officers to ensure that their firms operate in compliance with the legislation.²⁰⁵ The Standing Committee for Delegated Legislation specifically addressed the matter of criminal liability for directors and company officers during their discussion of the draft Water Order. Mr John McFall (the Parliamentary Under-Secretary of State for Northern Ireland) provided the Government's rather thin justification for denying the aquatic environment in this jurisdiction this potent form of protection. He explained that because individual liability for corporate offences "has been a thorny [problem] over the years", the Government in Northern Ireland "would prefer to learn from the experience of others on this matter."²⁰⁶

Anticipating the Adoption of the Water Framework Directive

As already stated, the EC is currently preparing to adopt a new Water Framework Directive.²⁰⁷ The purpose of the Directive is to incorporate all requirements for the management of water status into one single document and thus replace the 'first wave' of Directives on water quality adopted during the 1970s and early 1980s.²⁰⁸ Although it has been the subject of considerable controversy,²⁰⁹ it is generally accepted that the Framework Directive represents a radical re-thinking of European policy on the aquatic environment. The making of the Water (NI) Order 1999 provided the Government with an ideal opportunity to introduce the statutory basis for implementing the central elements of the Directive; regrettably, this opportunity was missed. Manifestation of this failure has already been discussed above in relation to the obligation to achieve water quality objectives; however, other notable failures to anticipate the requirements of the Directive should also be emphasised. The first striking example occurs in relation to the provision under the Order for making water management programmes. The core mechanism for ensuring the objectives set out in the Directive is the River Basin Management Plan (RBMP), identified by the Commission as the best model for a single system of water management. The Directive will require Governments to integrate a series of specific factors into each management plan; namely, the environmental objectives of the plan, an analysis of the characteristics of the river basin, a review of the environmental impact of human activity, an economic analysis of water use within the river basin, and the results of monitoring programmes carried out. Despite the fact that the Water Framework Directive will require Member

²⁰⁵ For a discussion of the liability of individuals for corporate offences see: Waite, "Criminal liability of Company Directors" (1991) *Land Management and Environmental Law Report*, 74; and Addison and Mack, "Creating an Environmental Ethic in Corporate America: The Big Stick of Jail Time" (1991) 44 *Southwestern Law Journal* 1427.

²⁰⁶ Parliamentary Debates, House of Commons: 2 March 1999, p 18.

²⁰⁷ Note 146 *supra*.

²⁰⁸ Commission Texts, note 146 *supra*.

²⁰⁹ Howarth, note 160 *supra*; Staallworthy, 'Water Quality: Proposal for a Draft Water Framework Directive' (1998) 9 *Water Law* 127; Jordan, "European Community Water Policy Standards; Locked in or Watered Down?", (1999) 37 *J Common Mkt Studies* 13.

States to ensure that an integrated RBMP is produced within each river basin, article 57 of the Water (NI) Order 1999 essentially repeats the entirely discretionary powers conferred on the DoE(NI) under the Water Act (NI) 1972 to make water management plans.²¹⁰ It is submitted that in light of the forthcoming adoption of the Framework Directive, the Water Order should have placed the Department under an obligation to make integrated water management plans consistent with the approach adopted by the Directive. A similar approach is taken in relation to the statutory controls imposed on the abstraction and impounding of water. Although the Water Order confers new powers on the DoE(NI) to adopt a system of licensing for the purpose of controlling the abstraction and impounding of water (including the power to establish a system of licence fees), the Department is not placed under an obligation to exercise these powers. The proposed Water Framework Directive will, however, require Member States to introduce controls over the abstraction of fresh surface water and groundwater, including a register of water abstractions and a requirement of prior authorisation for abstraction.²¹¹ These mandatory controls will also extend to the impounding of water. Not only are the DoE(NI)'s discretionary powers inconsistent with the compulsory nature of the controls required in this context by the Directive, even the introduction of a statutory licensing system will be insufficiently comprehensive as a system of control to satisfy the requirements forthcoming in the Directive.²¹²

Ensuring Effective Regulation in Practice

Throughout the two parts of this article, frequent reference has been made to the highly critical report published in April 1998 by the Comptroller and Auditor General for Northern Ireland concerning the control of river pollution in Northern Ireland.²¹³ It is not intended in this section to rehearse all of the specific findings made by the Auditor General. However, given the seminal nature of the report as the first detailed²¹⁴ and publicly available examination of the administration and enforcement of the Water Act (NI) 1972, it is important to emphasise the overall message conveyed by the Auditor General concerning this function. The central message emerging from the report is that the aquatic environment in this jurisdiction has not been adequately protected in the past. Virtually every aspect of the regulatory practice adopted by the DoE(NI)'s Environment and Heritage Service (EHS) was deemed to be deficient in fundamental respects. The

²¹⁰ Section 3.

²¹¹ Article 13.

²¹² See also the submission made by the RSPB(NI) during the consultation process surrounding the adoption of the Water Order – para 6.1-6.5. The RSPB(NI) also pointed out that the “the absence of comprehensive statutory controls over abstraction will have important implications for the delivery of sustainable water management in Northern Ireland. Even in areas with relatively plentiful water, abstraction control is still essential.”

²¹³ *Control of River Pollution in Northern Ireland*; HC 693, Session 1997-98.

²¹⁴ Although the legal protection of the aquatic environment was also addressed by the House of Commons Environment Select Committee Report in 1990 (HC 39, Session 1990-91, at paras. 72-80, pp xxiv-xxv) the Auditor General's report is the culmination of a six-year review which provides a much more detailed and specialised examination of this aspect of the EHS regulatory function.

Auditor General's specific findings concerning the administration and enforcement of the discharge consent system were particularly disquieting given the central role played by this particular mechanism in the regulation of water pollution in Northern Ireland. Four of the shortcomings identified in this regard are sufficiently serious to merit enumeration. First, the report revealed that although the EHS uses the Environment Agency's consent manual as a guide to setting trade consents, it has "no formal documentation of the operation of the consent-setting process in individual cases."²¹⁵ This lack of documentation was deemed to "preclude effective management of this function and may expose the EHS to criticism in the event of dispute over individual consent conditions."²¹⁶ Secondly, the EHS has only very recently (1997) acquired the technology necessary to measure the overall effectiveness of the consent system. The information produced by the new computerised information system indicated a lower level of compliance than in England and Wales.²¹⁷ Thirdly, the EHS has no formally stated procedures for enforcing discharge consents.²¹⁸ An examination of EHS action against non-compliant dischargers from 1990-95 indicated the its approach to enforcement "did not work in many cases."²¹⁹ Not only were dischargers allowed to remain in breach of their consents for "long periods of time before EHS takes any serious enforcement action",²²⁰ even where enforcement action is taken, the EHS approach to enforcement "does not achieve compliance within a reasonable timescale."²²¹ The EHS typically takes three months to issue warning letters²²² and even when issued, the Auditor General report described these letters as "surprisingly non-committal, often requiring no action or response from the discharger".²²³ The EHS reported to the Auditor General that the courts in Northern Ireland are "reluctant to convict in cases where an application for consent is outstanding."²²⁴ As a result, where the EHS attempts to prosecute, it has become "normal practice" to apply for a consent as a means of avoiding conviction.²²⁵ This problem was further compounded by the inordinate delays in issuing consents – instances of two and four-year delays were cited in the report.²²⁶

The Auditor General's report, which represents the culmination of a six-year examination of the EHS' work, raises very serious questions concerning its capacity to administer and enforce the more sophisticated controls introduced under the Water (NI) Order 1999 in an effective manner. Although it is essential that all of the Auditor General's recommendations are implemented, immediate improvement in certain areas is vital. The EHS must be given the resources necessary to develop an unambiguous and

²¹⁵ Note 213 *supra*, at p 43, para 3.5.

²¹⁶ *Ibid.*

²¹⁷ *Ibid.*, pp 44-46.

²¹⁸ *Ibid.*, p 72, para 7.2.

²¹⁹ *Ibid.*, p 76, para 7.14.

²²⁰ *Ibid.*, p 74, para 7.9.

²²¹ *Ibid.*, pp 72-73, para 7.4.

²²² *Ibid.*, p 74, para 7.8.

²²³ *Ibid.*, p 74, para 7.7.

²²⁴ *Ibid.*, p 75, para 7.11.

²²⁵ *Ibid.*, p 75, para 7.11.

²²⁶ *Ibid.*, p 75, para 7.12.

coherent enforcement policy under which all polluters are treated equally.²²⁷ EHS enforcement policy must also accord with the principles of enforcement established by the Environment Agency: namely, fairness, proportionality, transparency and consistency.²²⁸ The weaknesses in the monitoring of water quality control, discussed above, must also be addressed as a matter of urgency. Last, but not least, formal targets on compliance with discharge consents must be included as one of the EHS key performance targets with progress published in the EHS Annual Report. If the quality of freshwater in Northern Ireland is to be safeguarded from the potentially destructive effects of the pressure for economic regeneration it is crucial that the discharge consent system – as the central pivot in the regulatory framework – operates effectively.

CONCLUSION

It is clear from the foregoing analysis of the Water (NI) Order 1999 that its commencement has the potential to mark the beginning of a new era in the regulation of water pollution in Northern Ireland. However, if the promise of increased protection for the aquatic environment represented by the Order is to bear fruit in reality, then further reform is essential. The introduction of more sophisticated legislation on water pollution control is a pointless exercise if the body responsible for delivering pollution control is unable to administer and enforce its provisions in an effective and independent manner. Similarly, modern standards of pollution control cannot be guaranteed if the legal framework within which regulation takes place is weak in a number of fundamental respects in terms of the obligations it imposes on both the regulator and polluter alike. Although the Auditor General for Northern Ireland acknowledged that it was impossible to place a monetary value on improved water quality which could be set against the cost of achieving it, his recent report identified the value of fishing, both commercially and as a tourist amenity, as one possible indication of the importance of clean water for Northern Ireland.²²⁹ Figures provided by the Department of Agriculture (NI) estimated that commercial fishing on inland waters is worth £4.0 million to the local economy annually. Angling was considered to be worth £12.0 million annually and fish farming £3.3 million. The Auditor General pointed out that, in contrast, angling was worth £70 million to the Scottish economy annually and £35 million to the Irish economy.²³⁰ Although angling in Northern Ireland is currently underdeveloped as a tourist amenity, the Northern Ireland Tourist Board reported that it has “significant growth potential, and even at current levels, it is very significant to local tourist economies.”²³¹ The protection of the aquatic environment in Northern Ireland is now at a crossroads. If devolution is revived or at the very least, the paramilitary ceasefires hold, the Province is likely to experience a period of sustained and possibly rapid economic development. Northern Ireland’s inland waterways – with their traditional image of high environmental quality – are likely to figure prominently as a

²²⁷ For a discussion of EHS enforcement policy see, *ibid*, pp 68-71 and 73-76.

²²⁸ *Ibid*, p 68, para 6.13.

²²⁹ Note 213 *supra*, p 24, para 6.

²³⁰ *Ibid*.

²³¹ *Ibid*.

natural asset ripe for development as a tourist amenity. In addition, wider economic growth will be attended by greater risks of polluting discharges to water and an increased demand for clean water for both commercial and domestic uses. However, even if political violence does re-emerge, the need for modern standards of pollution control will remain, if for no other reason than because forthcoming EC water Directives will demand the delivery of more effective regulation. In order to safeguard the long-term future of one of Northern Ireland's most valuable economic and cultural assets, the Government must make a commitment to effective and proactive regulation, placing a heavy emphasis on the implementation of the dual principles of polluter pays and pollution prevention – regardless of progress towards wider constitutional settlement.

GRANT-ING RIGHTS: THE POLITICS OF RIGHTS, SEXUALITY AND EUROPEAN UNION

*Carl F. Stychin, Professor of Law and Social Theory,
University of Reading **

INTRODUCTION

On 17 February 1998, the European Court of Justice delivered its judgment in *Grant v South-West Trains*, holding that a refusal by an employer to grant travel concessions to a person of the same sex with whom a worker has a “stable relationship” does not constitute discrimination based directly on the sex of the worker, as prohibited by Article 141 (then 119) of the EC Treaty and by the Equal Pay Directive, even where such concessions are allowed to a person of the opposite sex with whom a worker has a stable relationship outside marriage.¹ The reaction to this decision by activist and academic commentators alike was both surprise and indignation: the latter generated by the apparent “injustice” of the decision on the basis of formal equality; the former, because the Advocate General had reached the contrary result in an opinion on 30 September 1997. For the legal rights-oriented activist group “Stonewall”, the outcome of this litigation strategy – which they actively supported – was somewhat embarrassing as well as unexpected, underscoring how the privileging of a rights based activist programme can lead to disappointing results. Not surprisingly, Stonewall assumed the moral high ground, describing the judgment as “a blow to lesbians and gay men everywhere in the EC”.² Academics have found the decision in *Grant* a rich source for analysis. The decision of the European Court of Justice has been deconstructed from a queer theoretical perspective,³ interrogated for what it reveals about the emerging conception of European citizenship,⁴ and examined for what it suggests about the relationship between the institutions of the European Union.⁵

* Earlier versions of this article were presented at a Department of Law Seminar, Queen’s University of Belfast (October 1999), at a “Law and Political Theory” Workshop, sponsored by the Department of Law, Keele University (November 1999), and at a School of Law Seminar, University of Nottingham (February 2000). Many thanks to the participants at those gatherings for their many comments and suggestions, and thanks to Jo Shaw and Davina Cooper for helpful comments on earlier versions.

¹ Case C-249/96 [1998] ECR I-621.

² *Stonewall News*, 17 February 1998 (<<http://www.stonewall.org.uk/stonewall/cgi-bin/news.cgi?equalpaylost>>).

³ Beger, “Queer Readings of Europe: Gender Identity, Sexual Orientation and the (Im)Potency of Rights Politics at the European Court of Justice” (2000) 9 *Social and Legal Studies* 250.

⁴ See *eg*, Bell, “Shifting Conceptions of Sexual Discrimination at the Court of Justice: from *P v S* to *Grant v SWT*” (1999) 5 *ELJ* 63.

⁵ See *eg*, Armstrong, “Tales of the Community: Sexual Orientation Discrimination and EC Law” (1998) 20 *Journal of Social Welfare and Family Law* 455.

In this article, I will also touch upon some of these themes, but the focus of analysis is rather different, in that I will deploy the *Grant* litigation in order to illuminate the wider limitations and distortions which rights politics can foster when it is privileged so centrally within an activist strategy. In particular, given the absence of any explicit basis for sexual orientation discrimination protection within EC (or UK) law, the turn to Article 141, I argue, was most probably a deeply flawed strategy from the outset; which, if successful, also would have distorted the meaning of this sex discrimination provision. I develop this (admittedly contentious) argument from the perspective of an emerging discourse on the political economy of rights, sexuality, and citizenship, which to date has been articulated primarily within North American debates on the politics of rights and sexuality.⁶ My claim is that given the focus of European rights discourse on the “market citizen”⁷ and the creation of a “transnational capitalist society”,⁸ the sometimes taken-for-granted politics of European sexual citizenship rights is much in need of interrogation for its broadly political economic implications. In so doing, the aim is to question the privileging of rights discourse in sexual citizenship strategies, particularly within the European arena, because of its limitations as a vehicle for challenging underlying structural barriers to full citizenship. I advocate instead, not a sceptical approach to European politics and law, but an engagement of activism with the construction of, and participation in, democratic institutions, and a politics, not only of legal recognition, but more broadly, of recognition *and* redistribution.

BACKGROUND TO THE DECISION

The facts and background to the *Grant* litigation have become extremely well known and thus may be summarised only briefly here. Lisa Grant, an employee of South-West Trains Ltd (SWT), was refused a travel pass for a female partner, despite the fact that her contract of employment stated that “you will be granted such free and reduced rate travel concessions as are applicable to a member of your grade. Your spouse and dependants will also be granted travel concessions”.⁹ It was company policy to grant “privilege tickets” to a common law spouse of the opposite sex provided a statutory declaration was made that a meaningful relationship had existed for a period of two or more years.¹⁰ Grant’s request for travel concessions was denied because her partner was of the same sex. She claimed that this was contrary to EC law given that her male predecessor in post had received the benefit for his opposite sex partner.

An industrial tribunal referred several questions to the European Court of Justice, all of which turned on the issue of whether SWT’s actions

⁶ See eg, Goldberg-Hiller, “‘Entitled to be Hostile’: Narrating the Political Economy of Civil Rights” (1998) 7 *Social and Legal Studies* 517; Boyd, “Family, Law and Sexuality: Feminist Engagements” (1998) 8 *Social and Legal Studies* 369.

⁷ Everson, “The Legacy of the Market Citizen” in Shaw and More (eds), *New Legal Dynamics of European Union* (1995), p 73.

⁸ Ball, “The Making of a Transnational Capitalist Society: The Court of Justice, Social Policy, and Individual Rights Under the European Community’s Legal Order” (1996) 37 *Harv Int’l LJ* 307.

⁹ [1998] ECR I-621 at 623.

¹⁰ *Ibid.*

constituted “discrimination based on sex” for the purposes of Article 141, the Equal Treatment Directive and the Equal Pay Directive. The case received the support of Stonewall, which “mounted an intense media campaign to raise awareness of the significance of a ruling favourable to Grant”.¹¹ Before the ECJ, both the United Kingdom and France intervened in support of SWT.

Despite a finding in favour of Grant by Advocate General Elmer, who reasoned that this difference in treatment constituted discrimination on the basis of “gender”,¹² the ECJ held that the refusal to grant travel concessions did not constitute discrimination based directly on the sex of the worker prohibited by Article 141 of the EC Treaty or the Equal Pay Directive.¹³ Nor did sex discrimination include discrimination on the basis of sexual orientation. First, the condition applied in the same way to male and female workers, and therefore could not be regarded as constituting discrimination directly based on sex.¹⁴ Moreover, discrimination based on sexual *orientation* did not constitute discrimination on the basis of the sex of a worker, and Community law does not regard stable relationships between two persons of the same sex as equivalent to stable relationships outside marriage between two persons of the opposite sex.¹⁵ According to the Court, “in those circumstances, it is for the legislature alone to adopt, if appropriate, measures which may affect that position”.¹⁶

Although mention was made by the Court of the International Covenant on Civil and Political Rights, and the view of the Human Rights Committee that “sex” “is taken as including sexual orientation”,¹⁷ that in itself could not give rise to a general principle of Community law:

“although respect for the fundamental rights which form an integral part of those general principles of law is a condition of the legality of Community acts, those rights cannot in themselves have the effect of extending the scope of the Treaty provisions beyond the competences of the Community”.¹⁸

In terms of institutional competence, the Court also noted that the Treaty of Amsterdam enables the Council of Europe – by unanimous vote on a proposal from the Commission after consultation with the European Parliament – “to take appropriate action to eliminate various forms of discrimination, including discrimination based on sexual orientation”.¹⁹

It is this judgment by the European Court of Justice, and the events which led to it, which form the backdrop against which my analysis proceeds.

¹¹ Armstrong, *supra* at 456.

¹² [1998] ECR I-621 at 629-630.

¹³ *Ibid.*, at 646.

¹⁴ *Ibid.*

¹⁵ *Ibid.*, at 648.

¹⁶ *Ibid.*

¹⁷ *Ibid.*, at 650, citing Communication No. 488/1992 *Toonen v Australia* (31 March 1994). I have considered the implications of the *Toonen* decision elsewhere; see Stychin, *A Nation by Rights: National Cultures, Sexual Identity Politics, and the Discourse of Rights* (1998), pp 145-193.

¹⁸ [1998] ECR I-621 at 650.

¹⁹ *Ibid.*, at 651.

THE POLITICS OF RIGHTS IN A EUROPEAN CONTEXT

The “rights debate” in legal theory is now sufficiently well known that it need not be replayed in this article.²⁰ The debate concerns the political implications of (primarily) constitutional rights struggles, particularly in a North American context, and it was propelled by a recognition of the inadequacy of liberal, modernist assumptions about the inherently progressive character of rights discourse. In reaction to that traditional legal “faith” in rights, some have argued that rights struggles tend to produce politically conservative (or more accurately) classic liberal outcomes, and that this provides the ideological underpinning of many rights “victories”.²¹ More progressive political struggles thus come to be channelled and neutralised through the turn to rights. Others are sceptical of this position, and point to the powerful emancipatory potential of the language of rights in some forms, and its role particularly in civil rights struggles in the United States.²² Still others see rights as having a highly disciplinary function, particularly when conjoined to the language of responsibility.²³ In this article, I reject the imperative to make a decisive choice as between these various positions. Rather, my approach might broadly be described in terms of the “critical pragmatism” advocated by Didi Herman, namely, that rights have their pragmatic uses depending upon the precise context, but that rights struggles should not be divorced from broader social, political and economic movements for progressive change.²⁴

In fact, the *Grant* litigation contains many diverse strands of the rights debate. Although in this article I focus primarily on the ideological grounding of rights which was articulated (explicitly and implicitly) in the case, and how potentially progressive rights discourse was largely “transformed into dominant ideological terms”,²⁵ I make no claim that this is the “grand narrative” of the litigation. Instead, it demonstrates simultaneously that rights can function importantly as a “heuristic device” for social movement politics, drawing attention to social struggles and, indeed, diverse “ways of life”.²⁶ It thus can educate people about political change, citizenship demands, and also – in the context of a European rights struggle – the potentiality and limitations of membership in a transnational community.²⁷

²⁰ For a good overview, see Herman, “Beyond the Rights Debate” (1993) 2 *Social and Legal Studies* 25.

²¹ See eg, Bakan, *Just Words: Constitutional Rights and Social Wrongs* (1997); Fudge and Glasbeek, “The Politics of Rights: A Politics with Little Class” (1992) 1 *Social and Legal Studies* 45.

²² See eg, Williams, “Alchemical Notes: Reconstructing Ideals from Deconstructed Rights” (1987) 22 *Harv Civil Rights–Civil Liberties L Rev* 401.

²³ See eg, Brown, *States of Injury: Power and Freedom in Late Modernity* (1995).

²⁴ Herman, “The Good, the Bad, and the Smugly: Perspectives on the Canadian Charter of Rights and Freedoms” (1994) 14 *OJLS* 589.

²⁵ Bakan, *supra* at 117.

²⁶ Kingdom, “Transforming Rights: Feminist Political Heuristics” (1996) 2 *Res Publica* 73.

²⁷ See Shaw, “The Politics of Membership in EU Citizenship” in Bankowski and Scott (eds), *The European Union and its Order* (2000), forthcoming.

Moreover, an analysis of the use of EC sex equality law as a tool in rights struggles importantly demands that the insights flowing from the academic commentary on the politics of rights be contextualised in the unique circumstances of European Union. Certainly, EC sex equality law illustrates both the practical potentiality and the limitations to the use of rights discourse. More broadly, the role of rights themselves in the EU legal order is specific, and should be differentiated from the rights debates that go on in North American academic communities. As Carlos Ball suggests, constitutional rights in the American sense are grounded in the value of individual autonomy, and rights are ostensibly designed to facilitate the individual atomistically choosing the ends that she desires to achieve.²⁸ By contrast, in the EU, freedoms and rights have had a more instrumental role, in that their fundamental character is a result of “their consequentialist function, namely their being necessary to achieve Community objectives”.²⁹ As a result, the protection of individual rights is geared, not towards a belief in the inviolability of individual autonomy, but rather, rights become a *means* “of enforcing the positive obligation of the Community and of the member states” towards the achievement of the objectives of the integration project.³⁰ The objectives, as is well known, are highly economic in character: “the creation of an integrated and ‘efficient’ market” underpinned by a “neo-liberal market ideology”.³¹

Thus, while some North American academics of the “left” go to great pains to demonstrate (often persuasively) the ideological grounding of rights discourse, those interested in European rights need not follow such a tortuous path. Individual rights have an *explicit* ideology, and while they can and often have been deployed towards moderately progressive ends – particularly in the context of employment related rights – such successes do not negate the claim that there is an ideological underpinning to rights discourse centred on market relations. Thus, for example, Tamara Hervey argues that a hierarchy exists among EU citizens with respect to social security, and at its apex “is the migrant EU citizen who is an employee or a self-employed person”.³²

The history of sex equality rights in EC law provides ample evidence for this argument. It is universally acknowledged in EC legal history that economic factors were the motivating force behind Article 141; namely, the perceived need to avoid distortions in competition between member states which had differing levels of protection for sex equality rights in the workplace.³³

²⁸ Ball, *supra* at 333.

²⁹ *Ibid.*, at 340.

³⁰ *Ibid.*, at 344.

³¹ Hervey, *European Social Law and Policy* (1998), p 204.

³² *Ibid.*, at 106.

³³ See *eg*, Barnard, “The Economic Objectives of Article 119”, in Hervey and O’Keeffe (eds), *Sex Equality Law in the European Union* (1996), p 321. Article 141 EC provides for equal pay for men and women, and has been supplemented by Directives concerning equal pay and equal value (75/117), equal treatment in employment (76/207), equal treatment in social security (79/7), equal treatment in occupational pensions (86/378), and equal treatment for self-employed women (86/613). For an outline of the legal framework, see Hervey and Shaw, “Women,

Subsequently, in the landmark case of *Defrenne 2*, the objectives of EC sex equality law were described as two fold, embracing both the need to avoid distortions in competition, and also the desire for social progress and the improvement in the working and living conditions of the peoples of Europe.³⁴

The effect of European sex equality rights in practice has been well documented, particularly by feminist analyses which have highlighted the positive benefits of the resort to rights, but also the limitations imposed both by the structure of the rights, and the broader ideological project of the formation of an internal *market* in which rights discourse is embedded.³⁵ The dominant focus on formal equality and equal opportunities in the workplace has had a “revolutionary affect”³⁶ for some women – namely, working European women, particularly around issues such as part time work and pregnancy discrimination. And it has been through rights struggles – often by “lone women” – that these successes have been realised as a result of often protracted litigation.³⁷

However, the limitations of EC sex equality law are also well known. The emphasis on “fair” competition in the marketplace, the “merit” principle, and “equal” opportunities leaves little scope for the use of rights discourse to tackle the underlying structural barriers to substantive equality, many of which result from the private sphere of the home and from impediments to full entry into the labour market, such as “the double burden of ‘care’ and ‘work’ for women”.³⁸ This is a realm considered beyond the role of rights which, because of the ideological basis of those rights, is focused on the public sphere, employment relation.³⁹ There is no recognition, for example,

Work and Care: Women’s Dual Role and Double Burden in EC Sex Equality Law” (1998) 8 *Journal of European Social Policy* 43 at 46-47.

³⁴ Case 43/75 *Defrenne v Sabena (No. 2)* [1976] ECR 455.

³⁵ See *eg*, Barnard, *supra*; McGlynn, “EC Sex Equality Law: Towards a Human Rights Foundation” in Hervey and O’Keeffe, *supra* at 239; Everson, *supra*; Hoskyns, *Integrating Gender: Women, Law and Politics in the European Union* (1996); More, “Equality of Treatment in European Community Law: the Limits of Market Equality” in Bottomley (ed), *Feminist Perspectives on the Foundational Subjects of Law* (1996), p 261; Hervey and Shaw, *supra*.

³⁶ Everson, *supra* at 209.

³⁷ Hoskyns, *supra* at 78.

³⁸ Hervey and Shaw, *supra* at 60. See generally Fredman, *Women and the Law* (1997). Of course, it is not simply EC sex equality law which can be criticised for its inability to deal with structural disadvantage. Rather, it is a critique which can be levelled more broadly at liberal anti-discrimination legislation: see Fredman, *supra*; Lacey, “Legislation Against Sex Discrimination: Questions from a Feminist Perspective” (1987) 14 *JLS* 411; McColgan, *Just Wages for Women* (1997). Recent European legislative initiatives, to some extent, can be interpreted as tackling that structural disadvantage: *eg*, the Parental Leave Directive, the Pregnancy Directive, and Child Care initiatives.

³⁹ See *eg*, Case 184/83 *Ulrich Hofmann v Barmer Ersatzkasse* [1984] ECJ 3047 at 3075: “It is apparent. . . that the directive is not designed to settle questions concerned with the organization of the family, or to alter the division of responsibility between parents”.

of voluntary work and informal care as leading to entitlements to rights such as social security.⁴⁰

Thus, two of the many sides of rights discourse become apparent. The language of rights has meant that “the EU system can be politicised in the interests of the democratic majority”⁴¹ (and, in that sense, rights prove a useful heuristic device), but, by virtue also of the explicit ideological grounding of EC law, rights struggles are channelled into an economically liberal model tied to the atomistic individual actor freely and fairly working in the competitive labour market.⁴² The potential for social change through the employment relation is certainly present (and has been achieved to some degree), but the role of rights in social change is constrained from the outset.

It is on such a politically ambiguous terrain that actors engaged with sexual orientation struggles have sought to “graft” their claims. Such a move is practically and politically problematic, however, in several important respects. Most obviously, such a move runs into difficult questions regarding the interpretation of the words of the provision, and the fairly clear intention that Article 141 was not intended to cover sexual orientation discrimination. Of course, it also can be argued that if the Treaty is a central constitutional document, then it should be construed in a purposive and teleological fashion and, if Article 141 is designed to foster social progress, then a broad interpretation is justifiable.⁴³ The issue, according to Kenneth Armstrong, is how two central tensions in EC law are to be resolved:

“The first is the extent to which the ECJ is willing to extend the scope of Community law beyond the domain of economic integration and to embrace the broader political dimension of laying the legal foundations for a citizenship of the EU. The second . . . is the tension between the interpretation and construction of the EC Treaty as a typical agreement between nation-states or as a constitutional text to be given meaning in the context of a process of constitution building”.⁴⁴

These core tensions suggest that a victory in *Grant* was never going to be straightforward or inevitable.

As well, the question of institutional legitimacy for the European Court of Justice inevitably pushed it away from such a broad interpretation of Article 141.⁴⁵ The inclusion of Article 13 EC as a result of the Treaty of Amsterdam may suggest that this is an area for legislative, as opposed to judicial,

⁴⁰ Ackers, “Citizenship, Gender and Dependency in the European Union: Women and Internal Migration” in Hervey and O’Keeffe, *supra* at 221, 226.

⁴¹ Hoskyns, *supra* at 210.

⁴² See Shaw, “Law, Gender and the Internal Market” in Hervey and O’Keeffe, *supra* at 283. The role of rights is undoubtedly multifaceted, and my focus is undeniably narrow. As well, rights should not be singled out as the sole means by which the European Court of Justice pursues legal integration; see Hilson and Downes, “Making Sense of Rights: Community Rights in E.C. Law” (1999) 24 *ELR* 121.

⁴³ Armstrong, *supra* at 462.

⁴⁴ *Ibid.*

⁴⁵ On the problem of legitimacy and the institutions of the European Union, see generally de Búrca, “The Quest for Legitimacy in the European Union” (1996) 59 *MLR* 349.

activity. Although it remains open to speculation whether and how this enabling treaty article will be implemented in the immediate (or, indeed, longer term) future, that in itself may be a reason for the Court to exercise self-restraint. As Steve Terrett suggests, “at a time when ratification of the Amsterdam Treaty [wa]s by no means a certainty, the ECJ may have felt it prudent to refrain from providing ammunition to eurosceptics in the various Member States by adopting a gung-ho approach to Community legal development”, an approach which might well have been perceived as illegitimate action on the part of the Court.⁴⁶ Indeed, the emotiveness of the combination of sexuality, rights and judicial activism has frequently resulted in claims of judicial illegitimacy in other constitutional jurisdictions and contexts.⁴⁷

A third problem is closely related to the institutional issue, namely, the difficulty of finding a level of uniformity in views across the member states sufficient to warrant judicial activism. Terrett analyses this point doctrinally in terms of the way in which the ECJ was asked to widen the meaning of “sex” discrimination on the basis of a general principle of law. As he argues, the essential requirement for recognising a general principle of law is “that the principle should be widely accepted by the Member States”, and that it “will require some level of uniformity, albeit short of precise consensus, before it [the Court] is willing to incorporate a principle into the EC system and offer it protection at a Community level”.⁴⁸ One of the explicit bases for the Court’s decision was the absence of such a consensus to ground a legal principle against discrimination on the basis of a person’s sexuality:

“As for the laws of the Member States, while in some of them cohabitation by two persons of the same sex is treated as equivalent to marriage, although not completely, in most of them it is treated as equivalent to a stable heterosexual relationship outside marriage only with respect to a limited number of rights, or else is not recognised in any particular way”.⁴⁹

This recognition of variation and difference in the attitudes of member states might well be understandable in terms of the relationship of sexuality to the private sphere. Again, in terms of legitimacy, Catherine Hoskyns argues that the ECJ has been of the view that “intervening in personal or domestic matters is not the function of either EC law or the Court”.⁵⁰ Rather, the personal becomes associated with the national, even though the precise issue – right to equal pay – is quintessentially a public, Community law matter.⁵¹ The alternative approach in these circumstances would be the judicial

⁴⁶ Terrett, “A Bridge too Far? Non-Discrimination and Homosexuality in European Community Law” (1998) 4 *European Public Law* 487 at 505.

⁴⁷ See generally my analysis in *A Nation by Rights*, *supra*.

⁴⁸ Terrett, *supra* at 498.

⁴⁹ [1998] ECR I-621 at 647.

⁵⁰ Hoskyns, *supra* at 160. See also Kofman, “Citizenship for Some but not for Others: Spaces of Citizenship in Contemporary Europe” (1995) 14 *Political Geography* 121 at 132.

⁵¹ On the public/private distinction in EC law, see generally More, “‘Equal Treatment’ of the Sexes in European Community Law: What does ‘Equal’ Mean?” (1993) 1 *Feminist Legal Studies* 45.

“recognition” (or, more accurately, imposition) of a principle of Community law. Arguably, that is the approach which the Court took in *P. v S. and Cornwall County Council*, in which it took no notice of national variation in the treatment of transsexuals.⁵² In interpreting that decision, Larry Backer suggests that judicial interpretation can act as a form of “normalizing harmonization”,⁵³ in which “subnational cultural determinism” gives way to the discipline imposed by legal standards.⁵⁴ That tension between harmonisation and self-determination is common in rights claims around sexuality⁵⁵ and, indeed, it can be argued that the politics of sexuality is characterised by a dialectical relationship between the local and the global.⁵⁶ This tension may well have been an important factor which motivated the Court to defer to the local, and it is a factor closely related to the issue of judicial legitimacy.

Finally, it has been argued that a central problem with this litigation strategy was, quite simply, “the facts”. As is well known from the history of civil rights struggles in the United States, constitutional litigation strategies demand compelling “test cases”, and Mark Bell has argued that the “justice” of the issue in *Grant* was simply not overwhelmingly compelling.⁵⁷ By contrast, a set of facts dealing with an outright, irrational dismissal from employment (as was the case in *P. v S.*) may well have resulted in a different outcome.⁵⁸ The granting of employment “perks” may well seem to many a less than compelling human rights case, particularly when those perks are not granted to employees who are not in any sort of traditional spousal-type relationship and who thus continue to suffer this “discrimination” no matter what the result of the case.

Although Bell’s point is intuitively appealing, I want to argue, by contrast, that the claim in *Grant* does fit nicely into the ideological parameters of EC law, and particularly European rights discourse, despite the fact that the claim was ultimately unsuccessful. In making this argument, I hope to illustrate again how law has an often complex and contradictory role in social movement politics, and this is exacerbated in the realm of EC law. Thus, although the facts of *Grant* may appear to lack the moral imperative of a human rights claim, that is also arguably why it fits within this paradigm of rights discourse. Indeed, it has been argued that the advantage of deploying EC law for gays and lesbians is that the economic paradigm of rights

⁵² [1996] IRLR 374. See also my analysis of the decision: Stychin, “Troubling Genders: A Comment on *P. v S. and Cornwall County Council*” (1997) 2 *Int’l Journal of Discrimination and the Law* 217.

⁵³ Backer, “Harmonization, Subsidiarity and Cultural Difference: An Essay on the Dynamics of Opposition Within Federative and International Legal Systems” (1997) 4 *Tulsa Journal of Comp and Int’l Law* 185 at 197.

⁵⁴ *Ibid.*, at 199.

⁵⁵ See generally Stychin, *A Nation by Rights*, *supra*.

⁵⁶ See Nardi, “The Globalization of the Gay and Lesbian Socio-Political Movement: Some Observations about Europe with a Focus on Italy” (1998) 41 *Sociological Perspectives* 567.

⁵⁷ Bell, *supra* at 78-79.

⁵⁸ In fact, it has been argued that the Court of Justice might still find in favour of a claimant in a sexual orientation employment recruitment or dismissal case; see Denys, “Homosexuality: a non-issue in Community law?” (1999) 24 *ELR* 419 at 425.

abstracts them from an obviously moral underpinning, making it easier to make claims in a morally “neutral”, economically grounded language.⁵⁹ As a consequence, “successes” will be more likely. Discrimination becomes a distortion of the transnational marketplace and a barrier to free movement, and the sort of controversies which are fuelled by gay rights litigation in other constitutional jurisdictions can be avoided. In other words, the economic teleology of rights in the EU can “sanitise” the claim, making it more likely that a court will conclude that it can legitimately find in the claimant’s favour. Although I have argued that such an instrumental approach to rights as an activist strategy is misguided,⁶⁰ that instrumentalism does capture something about the ideology of rights discourse, in some of its forms.

In fact, the focus in *Grant* on relationships also closely fits the ideology of the “family” as it has developed in EC law. Louise Ackers and Helen Stalford have examined how the family is conceived in EC law in the context of the free movement provisions, in which a series of social rights for the families of EU migrant workers has been recognised, providing equal treatment protection in matters of employment, pay and working conditions.⁶¹ The assumption made by the European Court of Justice is that there is a close correlation between the exercise of the right of free movement and the granting of free movement rights to family members.⁶² However, Ackers and Stalford emphasise that “the rationale for the Court’s incursion into areas of family policy in this area of Community law is based firmly on a conceptualisation of women and children as the non-productive appendages of male workers”.⁶³ Moreover, only a marital relationship can be used to underpin the claim as far as the dependent partner of an EU migrant worker is concerned.⁶⁴ Thus, a “breadwinner” model of “coupledom” is assumed, in which labour mobility depends upon the ease with which the worker can move the dependent spouse with “him” when he, as a factor of production, is more highly valued in another member state. The facts of *Grant* tap into that same ideology, in which perks are for dependents, in a model of family based upon a breadwinner, “family” wage earner. Thus, *Grant* exemplifies a well known litigation strategy, in which test cases draw upon fact situations which are constructed so as to replicate very traditional gendered relationship patterns, albeit with a same sex twist. Replication is assumed to be the path to litigation success.

This strategy is problematic in several respects, and a variety of commentators have highlighted the various reasons why. For example, in the context of *Grant* itself, Amy Elman has argued that “bartering over privileges, rather than demanding their eradication – represents a lack of

⁵⁹ Ball, *supra* at 387.

⁶⁰ See Stychin, *A Nation by Rights*, *supra* at 143.

⁶¹ Ackers and Stalford, “Children, Migration and Citizenship in the European Union: Intra-Community Mobility and the Status of Children in EC Law” (1999) 21 *Children and Youth Services Rev* 699.

⁶² *Ibid*, at 705. See also Hervey, *supra* at 106.

⁶³ Ackers and Stalford, *supra* at 702.

⁶⁴ *Ibid*, at 708.

political judgment and a lack of ambition".⁶⁵ She suggests that the litigation symbolises how lesbian and gay movements have now "begun embracing patriarchy's principal institution, the family".⁶⁶ A similar argument is made by Momim Rahman, who claims that these strategies endorse, rather than challenge, institutionalised heterosexuality as a model for human relationships, which is then replicated by same sex couples.⁶⁷ Both Elman and Rahman argue in favour of wider strategies for social and economic change, a point to which I return later in this article.

However, these arguments are clearly contentious, and the tensions within the politics of *Grant* perhaps best can be illustrated through the academic debate staged between social and cultural theorists, Nancy Fraser and Judith Butler.⁶⁸ This is not the place to restage that debate, but suffice it to say that the basic issue which divides these two theorists is the relationship between struggles around the recognition of "sexual orientation", and wider issues of political economy and economic transformation. In the debate, this has been framed in terms of the language which Fraser has developed regarding the relationship between a politics of recognition and a politics of redistribution.⁶⁹ Fraser's point is that sexuality struggles are *essentially* about a politics of recognition, rather than about issues of redistribution in political economic terms. The two are separate struggles.

Butler, in response, has questioned the dichotomy, and has asked pointedly, "why would a movement concerned to criticize and transform the ways in which sexuality is socially regulated not be understood as central to the functioning of political economy?"⁷⁰ That is, Butler claims that sexuality must be understood as part of the mode of production itself. She also presents examples to refute Fraser's claim that recognition and redistribution are necessarily separate. Most obviously, lesbians as (marginalised) women as a group are going to experience both a wage gap (an issue of economic distribution), and a lack of social recognition.⁷¹

The relationship between recognition and redistribution claims has now begun to be analysed explicitly in legal scholarship. Susan Boyd, for example, argues that while Fraser's sharp dichotomy is problematic for precisely the reason that she seems to forget that lesbian women are gendered, and gender is central to the mode of production, Butler's position is also troubling.⁷² As Boyd notes, similar to the critiques offered by Elman and Rahman, "it does not follow that legal recognition of non-normative

⁶⁵ Elman, "Familiar Orientations: Sex Discrimination, Same Sex Partners and Migration in European Law", paper presented to the American Political Science Association Annual Conference, Atlanta, September 1999.

⁶⁶ *Ibid.*

⁶⁷ Rahman, "Sexuality and Rights: Problematising Lesbian and Gay Politics" in Carver and Mottier (eds), *Politics of Sexuality: Identity, Gender, Citizenship* (1998), p 79 at 85.

⁶⁸ See Fraser, *Justice Interruptus* (1997); Butler, "Merely Cultural" (1998) 227 *New Left Review* 33; Fraser, "Heterosexism, Misrecognition and Capitalism: A Response to Judith Butler" (1997) 15 (52/53) *Social Text* 279.

⁶⁹ Fraser, *Justice Interruptus*, *supra* at 12.

⁷⁰ Butler, *supra* at 39.

⁷¹ *Ibid.*, at 41.

⁷² Boyd, *supra* at 375-376.

sexualities (for example, same-sex relationships) will necessarily, of itself, constitute a fundamental challenge to the capitalist mode of production.”⁷³ Intuitively, given the purpose of EC rights discourse to further a free market transnational capitalist economy, it would be surprising if legal recognition in the European context would amount to such a challenge here.

In fact, the critique which Boyd offers – which is situated in the context of Canadian equality rights discourse around same sex partnership rights – is very similar to the feminist critiques of EC sex equality law. The argument in both contexts is that rights do little, if anything, to alter underlying structures which produce gender inequality, such as the role of unpaid labour in the private sphere, and barriers to entry in the workplace. Boyd argues that the nuclear family has a material role in capitalist relations of production, through a sexual division of labour and the privatisation within the nuclear family of the social costs of reproduction and care.⁷⁴ It is this same model of the nuclear family which, Ackers and Stalford argue, is privileged within EC legal discourse.⁷⁵ The agenda which Boyd then advances is one in which activists and academics should pay greater attention to whether that gendered economy is challenged by lesbian and gay legal struggles, or alternatively, whether the lesbian or gay subject is naturalised within the political economy through the (eventually successful) claiming of rights.⁷⁶ In this regard, it has been argued that the construction of sexual identities (and political priorities for a movement) historically has been shaped by the more privileged (in terms of social class) members of the group.⁷⁷ Thus, the question whether sexuality is integral to capitalism cannot be separated from whether sexual identities are significantly constituted and experienced in ways which reflect individual location within that mode of production.⁷⁸

Taking up Boyd’s challenge to interrogate rights claims, my argument is that the *Grant* litigation provides a perfect example of such a normalisation, even though the litigation ultimately proved unsuccessful. On the immediate facts, a successful claim in *Grant* would have seen the extension of marital-type perks to “stable” same sex couples, which presumably would have marginally increased employer costs. The actual “take up” of the marital privileges no doubt would vary greatly across the EU depending upon local and corporate social attitudes towards homosexuality, the prevalence of such benefits in the cultural context in issue, and, indeed, whether there was already anti-discrimination legislation in place at national level in the EU member state which covered this field. The wider procedural right of non-discrimination in employment in itself can be seen as a means to “perfect” the marketplace to the extent that anti-discrimination law is effective in

⁷³ *Ibid.*, at 376.

⁷⁴ *Ibid.*, at 376-377.

⁷⁵ Ackers and Stalford, *supra* at 709.

⁷⁶ Boyd, *supra* at 381.

⁷⁷ See eg, Valocchi, “The Class-Inflected Nature of Gay Identity” (1999) 46 *Social Problems* 207; Robson, *Sappho Goes to Law School* (1998), ch 12.

⁷⁸ As Iris Marion Young notes, “political economy is cultural, and culture is economic”: Young, “Unruly Categories: A Critique of Nancy Fraser’s Dual Systems Theory” (1997) 222 *New Left Review* 147 at 154.

eradicating the use of irrelevant characteristics in hiring, promotion and dismissal.

However, what is also significant, as Bell points out, is that the UK government submission in the *Grant* litigation focused on the wider implications and potential future litigation which might flow from a decision in favour of Grant.⁷⁹ In particular, state pension and social security issues may well have been at the forefront of concern. These are areas that are expensive for the state in a capitalist economy. After all, maintaining the attractiveness of the institution of marriage and marriage-like relationships requires costly social engineering. But such institutions are necessary for the maintenance of a capitalist system with a clear public/private dichotomy, in which many costs are internalised within the domestic sphere of the home.⁸⁰ Although the European Court of Justice often has been prepared to reject the appropriateness of economic justifications for sex discriminatory employment practices,⁸¹ many of these decisions are themselves ideologically “loaded” (such as those concerning protective treatment of pregnant workers), and do not in themselves refute the claim of an ideological basis to the legal order centred on the public/private dichotomy.⁸²

Thus, a decision in favour of Lisa Grant would have fitted very well with the ideological grounding of EC rights discourse, in terms of the nuclear family and privatised responsibility within that private sphere and it would have furthered transnational cultural harmonisation through a common EU definition of “spouse”. However, the wider implications of the decision – with the resulting social costs – would not be easily distinguishable within the “all or nothing” paradigm of formal equality in EC law. The economic costs of legally normalising the homosexual subject are greater in terms of numbers than those of normalising the transsexual subject in *P. v S*. The scale of such costs no doubt is an inhibiting factor for the judiciary. By way of contrast, Canadian courts have found it much easier to make such distinctions.⁸³ As a result, as Claire Young argues, we now see Canadian courts developing broad definitions of “spouse” in the context of upholding private obligations (such as support), while constructing narrow definitions in the context of state pensions.⁸⁴ These results in Canadian jurisprudence should also make us sceptical as to whether rights in the North American context are so far removed from the economic teleology of the EC legal

⁷⁹ Bell, *supra* at 76.

⁸⁰ Boyd, *supra* at 377. See also Hervey and Shaw, *supra*.

⁸¹ See *eg*, Case C-66/96 *Handels-og Kontorfunktionaerernes Forbund i Danmark* [1998] ECR I-7327 in the context of pregnancy. But see also Case C-394/96 *Brown v Rentokil* [1998] ECR I-4185 and Case C-179/88 *Handels-og Kontorfunktionaerernes Forbund i Danmark v Dansk Arbejdsgiver forening* [1990] ECR I-3979 (determining that pregnancy-related illness beyond the period of maternity leave is to be treated as sex-neutral).

⁸² See Mancini and O’Leary, “The New Frontiers of Sex Equality Law in the European Union” (1999) 24 *ELR* 331.

⁸³ Compare, *eg*, *M. v H.* [1999] 2 SCR 3 (extending definition of “spouse” for purposes of spousal support law) and *Egan v Canada* [1995] 2 SCR 513 (no extended definition in the context of old age security legislation).

⁸⁴ Young, “Spousal Status, Pension Benefits and Tax: *Rosenberg v Canada*” (1998) 6 *Cdn Labour and Employment L J* 435.

system. Had the European Court of Justice seen a way to achieve this distinction, a different result might have been forthcoming.

I make this hypothesis because, on its facts, *Grant* is ideologically very attractive in terms of the underpinnings of EC law: no immediate cost to the state; recognisable model of relationships suggesting economic dependence on a breadwinner; and the cost of the perk is a product of a contractually regulated private employment relationship. It provides a clear example of “the domestication of deviant sexualities within a safe, useful and recognisable framework”,⁸⁵ while the cost of the relationship does not touch the state. At the same time, a standardised, normalised definition of “spouse” modelled on heterosexual marriage creates a level playing field across the EU (from which deviant “others” can then be excluded).

A central paradox in the use of legal discourse towards the recognition of same sex relationships thus becomes apparent. Social scientists increasingly are confirming through empirically based research what many have long known: that lesbians and gays construct an infinite variety of ways of living – and of relationships widely defined – which not only replicate but also resist the disciplinarity of heterosexual monogamous cohabitation.⁸⁶ Yet, when legal discourse is deployed in activist struggles for social change, the engagement with law seems to require constructing relationships which replicate monogamous, heterosexual cohabitation as an “ideal” to which lesbians and gays can successfully aspire, and from which benefits then flow. These two dynamics – of resistance and discipline – appear to coexist simultaneously, although it is also important to avoid the privileging of legal discourse in analysing processes of social change.

While the successful deployment of rights strategies will benefit some lesbians and gays materially, and no doubt symbolically, in terms of the affirmation of relationships (for otherwise, there would never have been litigation), the underlying economic question remains: what is the relationship between such claims – framed within the context of a sex discrimination provision – to underlying gender-based structural inequalities? More specifically, is there a necessary relationship between claims to recognition and what Butler refers to as the “holy family” of the capitalist mode of production?⁸⁷

First, in engaging in such an analysis it is important to avoid the extreme positions which the Fraser-Butler debate produces. The particularities of any struggle are important, as is the cultural and political context. For example, relationship recognition struggles in the United States arise in a political context in which, for many, affordable health care is *dependent* upon establishing a relationship with someone who has a “good company plan”. Particularly in the context of new HIV therapies, this is a pressing issue, in which recognition issues and the material conditions of life are inextricably linked. Of course, at the same time, “success” through legal recognition is problematic to the extent that it detracts attention from the urgent need for

⁸⁵ Boyd, *supra* at 377.

⁸⁶ Donovan, Heaphy and Weeks, “Citizenship and Same Sex Relationships” (1999) 28 *Journal of Social Policy* 689.

⁸⁷ Butler, *supra* at 41.

universally available health care. Moreover, even in this example, rights can be politically and materially indeterminate, for legal recognition may well result in a financial privatisation of care responsibilities, onto the responsible “spouse”.⁸⁸ My general point is that the issue of health care benefits has a particularly American resonance which does not travel well, for example, to the European arena.

However, it does seem a reasonably general proposition that employment rights protection will have a differential impact depending upon the intersection of identities implicated in any individual case, and this point is also related to the distinction between formal and substantive equality. At this juncture, it is useful to recognise that rights discourse around sexual orientation will benefit *most* those for whom there are no other structural, identity based impediments to the realisation of substantive equality; or, in other words, where there is no intersection of disadvantage.⁸⁹ But there is another dynamic at work here, for all gay men (except to the extent that they are perceived to embody cross-gender characteristics) *already* benefit “from the institutions and social customs that hinder female entry into certain jobs, including traditional notions that women should be committed to domestic life and not to market labour”.⁹⁰ That is, to the extent that rights discourse can eliminate irrational anti-gay animus in the workplace (and I do not want to exaggerate the extent to which law can achieve this aim), gay, childless men can benefit tremendously from patriarchal economic relations which largely privatise responsibility for the care of the young.

Meanwhile, many lesbian women *qua* women may well still face the structural barriers which are not rectifiable through a formal equality rights based discourse of equal opportunities alone (although statistically as a group they may fare better in some economic respects than heterosexual women). Michael Jacobs neatly summarises a political dynamic which increasingly rings true today:

“Gay families need not be modeled on the social norm of a two-adult household, but for gay male couples who do form a two-income household, the benefit of men’s generally higher wage earnings is doubled. This amplifies income differentials by gender within the gay community and perhaps makes gay men more conscious of the economic benefits of male privilege and less comfortable with an agenda that challenges this privilege. To the extent that their efforts to integrate into their families of origin are successful as well, many gay men may no longer find the social dominance of the family

⁸⁸ See Flanagan, “People with HIV/AIDS, Gay Men, and Lesbians: Shifting Identities, Shifting Priorities” in Herman and Stychin (eds), *Legal Inversions: Lesbians, Gay Men, and the Politics of Law* (1995), p 195 at 210.

⁸⁹ I have developed a related argument concerning “intersectionality” in a different context; see Stychin, “Essential Rights and Contested Identities: Sexual Orientation and Equality Rights Jurisprudence in Canada” (1995) 8 *Cdn Journal of Law and Juris* 49.

⁹⁰ Jacobs, “Do Gay Men have a Stake in Male Privilege?: The Political Economy of Gay Men’s Contradictory Relationship to Feminism” in Gluckman and Reed (eds), *Homo Economics: Capitalism, Community, and Lesbian and Gay Life* (1997), p 165 at 170.

particularly oppressive and thus may not respond favorably to a political rhetoric that describes it that way".⁹¹

In addition, for those gay male couples who do not reap the benefits of a two-income household, employment perks such as those at issue in *Grant* at least will allow them to keep pace with the heterosexual single income household. The point which this analysis drives at is that the use of sex equality laws – even if successful on its own terms – will definitely benefit some more than others (ironically, men more than women), and must be considered at best a modest part of a wider strategy of social change.

Furthermore, Jacobs' point is that many gay men may well not have a tremendous stake in a wider strategy for social and economic change, particularly if they achieve the sort of formal legal guarantees which were at issue in *Grant*. Here again, the ideology underpinning the EU integration project illustrates how the construct of citizenship – which has been widely critiqued as a limited, market-centred concept by progressive legal academics⁹² – is also one which is actually advantageous in its current form to many gay men. In fact, the subject position of EU citizenship has been described as "archetypically male" because of its market centredness, in which the citizen is imagined enjoying the benefits of free movement unconstrained, allowing him to sell his labour to the highest bidder transnationally, and enjoying ever widening consumer choice to satisfy his wants and desires.⁹³ This citizen is highly atomistic, unconstrained by relationships of dependence, and whose primary identity is derived from paid employment. For many gay men, such a construct is very appealing, particularly if formal legal protection against discrimination ensures that they can exercise these citizenship rights without fear of irrational prejudice by others. Homosexuality then becomes simply, as Mariana Valverde suggests, a "lifestyle choice" and "an innocuous feature of urban consumer life" for the market citizen who can choose to consume homosexuality unconstrained – and across national borders.⁹⁴ Many gay men, in this citizenship discourse, are themselves archetypically male: more man than most ("new") heterosexual men who are increasingly expected to prioritise their familial relationships, and to share (albeit certainly not yet equally) in child care responsibilities. Thus, European law and social policy have been subject to critique for their "commodification of individuals" into workers (that is, producers).⁹⁵ But homosexuality is also increasingly criticised for its own form of commodification, centred on the consumption of goods, services and, indeed, a way of life, in which "consumerism becomes the embodiment of identity".⁹⁶

Of course, this vision of the gay male EU citizen is also dependent upon the marginalisation of traditional familial and religious discourses of

⁹¹ *Ibid*, at 172.

⁹² See generally the contributors to Shaw and More, *supra*.

⁹³ Shaw, *supra* at 297.

⁹⁴ Valverde, "'The Lifestyle that Fits the Doctrine of Sexual Orientation': The Disappearing Homosexual of Human Rights Law", paper presented to the American Law and Society Association Annual Conference, Chicago, May 1999.

⁹⁵ See Hervey, *supra* at 204.

⁹⁶ Valocchi, *supra* at 220.

homophobia; but, here again, the underlying ideology of EC law is sympathetic to such progressive change. A modernist discourse of capitalism has little room for such irrationality, which can serve to distort competitive labour markets, and an underlying desire for cultural harmonisation demands that one should be able to cross borders freely both to produce and consume a lifestyle unconstrained by bigotry in some member states. Historically, some lesbians and probably many more gay men have often benefited from capitalism, as the “marketplace cleared away all sorts of traditional social formations”.⁹⁷ As John D’Emilio has argued, it was only when wage labour became the widespread basis by which people lived that homosexuality could become the foundation of a personal identity.⁹⁸ Thus, for example, Peter Nardi makes the claim that in northern Italy, the preconditions for a lesbian and gay movement occurred only relatively recently, with the growth of urban middle class employment opportunities, economic development and personal mobility and the growth of non-religious associations.⁹⁹ These processes no doubt will continue and accelerate throughout the EU, particularly with the likely widening of its borders eastward.

The argument underscores that while capitalism has depended upon the heterosexual nuclear family to help maintain it, late capitalism can now accommodate (quite happily) a group of autonomous, unconstrained producers and consumers operating outside the traditional constraints of the nuclear family. One need only look to the “positive” attitude of many multinationals to gay consumers today for evidence of this point. It is somewhat ironic then that within legal discourse around relationships, activists and litigants seek to discipline citizenship into something which perhaps is inevitably going to replicate the nuclear “holy family”, at the same time as that “holy family” is increasingly decentred in many ways.¹⁰⁰ More than anything, this irony demonstrates how legal discourse functions and constrains argumentation through the demands of precedent and comparison. Finally, it suggests that Butler may underestimate the inherent conservatism of law when she suggests that same sex recognition struggles inevitably challenge the “holy family”, for it is only by making one’s self (or one’s client) look like the (admittedly decentred) norm that “success” in instrumental terms is likely to be achieved.

⁹⁷ Adam, Duyvendak and Krouwel, “Gay and Lesbian Movements beyond Borders?” in Adam, Duyvendak and Krouwel (eds), *The Global Emergence of Gay and Lesbian Politics* (1999), p 344 at 356.

⁹⁸ D’Emilio, “Capitalism and Gay Identity” in Snitow, Stansell and Thompson (eds), *Powers of Desire: The Politics of Sexuality* (1983), p 100.

⁹⁹ Nardi, *supra* at 579.

¹⁰⁰ In examining the Danish experience – “the first country in the world” to enact same-sex registered partnership legislation (*infra* at 234) – Karin Lützen points out that marital-type benefits have come to same sex couples in a society in which 25% of couples with children are unmarried; see Lützen, “Gay and Lesbian Politics: Assimilation or Subversion: A Danish Perspective” (1998) 35 *Journal of Homosexuality* 233 at 239. Thus, she concludes that it is only when marriage as an institution becomes anachronistic – “a worn-out piece of folklore” (*ibid*, at 239) – that its benefits are made available to homosexuals.

CONCLUSION: FROM EQUALITY TO ACTIVE CITIZENSHIP

Given what I have suggested about the relationship between capitalism and sexuality in the face of the decline of traditional social formations, it must be tempting (at least for those gays and lesbians who are secure in their jobs), to await the “liberalising effect of time”,¹⁰¹ particularly given the limitations inherent in the use of rights discourse to advance social change. Why not “opt out” of law, once basic privacy rights are recognised and a libertarian approach is accepted to the criminal law regulation of sexuality? In answer, it is worth remembering that while we may point to the disciplinary effect of rights, and the way in which progressive politics can be ideologically channelled into principles such as formal equality, *particularly* in a legal context such as the EU, the potentialities of rights discourse also must be recalled. The problem seems to be that although it may be recognised widely that rights through law provide incomplete political strategies, the problem with rights discourse is the way in which it seduces its users to believe in its totalising potential as a political strategy.

This is particularly true in the EU context. Given the democratic deficit, the democratic potentiality of rights discourse is necessarily constrained. Thus, on the one hand, we might applaud the pleas of commentators Philip Alston and Joseph Weiler that the EU needs a general equal treatment provision and draft directive on sexual orientation discrimination¹⁰² (which also has been called for by the European Commissioner responsible for Social Affairs and Employment¹⁰³). Such a provision – however politically unlikely it may seem – undoubtedly would have important practical and symbolic importance. However, Alston and Weiler may well underestimate the resistance which such a provision might meet.¹⁰⁴ Here again, the dialectical relationship between the local and the global must be recalled. In other national contexts, rights claims around sexuality have a unique power to fuel a reaction grounded in the discourses of localism and cultural and legal self-determination.¹⁰⁵ Moreover, as empirical research continues to demonstrate, “there is no monolithic Continental attitude towards sexuality in Europe”,¹⁰⁶ and, in particular, “homosexuality generates more varied opinions across countries and more polarized responses within nations” than do other contentious issues surrounding sexuality (such as extra-marital heterosexuality).¹⁰⁷

The other problem in the EU context, as Jo Shaw suggests, is that rights remain very much a “top-down” process, leaving little discursive space for

¹⁰¹ Scott, “Changing Attitudes to Sexual Morality: A Cross-National Comparison” (1998) 32 *Sociology* 815 at 839.

¹⁰² Alston and Weiler, “An ‘Ever Closer Union’ in Need of a Human Rights Policy” (1998) 9 *European Journal of Int’l Law* 658 at 717.

¹⁰³ *Stonewall News*, 7 December 1998 (<<http://www.stonewall.org.uk/stonewall/cgi-bin/news.cgi?employmentcommission>>).

¹⁰⁴ See generally Backer, *supra*.

¹⁰⁵ For example, Tasmania, Alberta, Colorado, and Zimbabwe.

¹⁰⁶ Widmer, Treas, and Newcomb, “Attitudes Toward Nonmarital Sex in 24 Countries” (1998) 35 *Journal of Sex Research* 349 at 356.

¹⁰⁷ *Ibid.*, at 352.

the power of rights as an enabling device for social movement activism.¹⁰⁸ Even in Alston and Weiler's argument, rights remain "granted" from on high, rather than being perceived as the product of years of social movement mobilisation.¹⁰⁹ Although I have argued elsewhere that in recent years there has been some space for NGOs such as the International Lesbian and Gay Association Europe to intervene in European rights debates around sexuality, it remains a modest space, because European identities and transnational communities are still very much in their infancy.¹¹⁰ However, as transnational communities, identities and affinities develop, there is no inherent reason why the European legal arena might not be one in which mobilisation and struggle from below could occur more widely.

In fact, one of the roles which sexuality politics might play is in seeking to expand that space of transnational politics, in an attempt to develop strategies of citizenship which go *beyond* the market towards a political and social citizenship of the European Union. The EU remains an important site, and there is some evidence that a notion of "social citizenship" has increasing currency through, for example, EU funding and "soft law",¹¹¹ as well as in some recent rights based discourse from the ECJ.¹¹² It is an arena which should not, then, be seen as solely one for the pursuit of legal rights, which can then be "consumed", but instead, as having the potential to mean something "more". Indeed, NGOs such as ILGA Europe do seek to combine calls for rights and democratic participation in the EU, and this provides a useful antidote to the privileging of rights discourse.¹¹³ It might be helpful here to imagine a floor of anti-discrimination rights – perhaps guaranteed through EU law – which acts as a base for the development of strategies of democratic participation within the institutions of the EU (and within social movements themselves).¹¹⁴

Richard Bellamy has argued that European rights will only be made meaningful in the context of democratic political arrangements.¹¹⁵ Otherwise, "rights prove too indeterminate and subject to conflicting interpretations to provide a constitutional basis for a European polity".¹¹⁶ In the EU, things are "further complicated by the existence of a plurality of national political

¹⁰⁸ Shaw, "European Union Citizenship: the IGC and Beyond" (1997) 3 *European Public Law* 413 at 427.

¹⁰⁹ Alston and Weiler, *supra*.

¹¹⁰ See Stychin, "Disintegrating Sexuality: Citizenship and the European Union" in Bellamy and Warleigh (eds), *Citizenship and Governance in the European Union* (2000), forthcoming.

¹¹¹ See Hervey, *supra* at 205.

¹¹² See, in particular, *Martinez Sala v Freistaat Bayern*, Case C-85/96 [1998] ECR I-2708.

¹¹³ See International Lesbian and Gay Association Europe, *Equality for Lesbians and Gay Men: A Relevant Issue in the Civil and Social Dialogue* (1998).

¹¹⁴ See Bellamy and Castiglione, "The Communitarian Ghost in the Cosmopolitan Machine: Constitutionalism, Democracy and the Reconfiguration of Politics in the New Europe" in Bellamy (ed), *Constitutionalism, Democracy and Sovereignty* (1996), p 111 at 122.

¹¹⁵ Bellamy, "The Constitution of Europe: Rights or Democracy?" in Bellamy, Bufacchi and Castiglione (eds), *Democracy and Constitutional Culture in the Union of Europe* (1995), p 153.

¹¹⁶ *Ibid.*

traditions”,¹¹⁷ and sexuality is certainly not exempt from these dynamics. Shared democratic arrangements are therefore necessary for the creation of common rights. This, I conclude, is a central limitation of litigation strategies such as *Grant*. Rights have been abstracted from any form of democratic politics.¹¹⁸ At the same time, rights claims can appear undemocratic from the perspective of those who are members of the social group at the epicentre of the claim, to the extent that the case may not reflect what are thought to be grassroots priorities by some activists; and, furthermore, rights may privilege the interests of some in the group over others.

Yet, at the same time, the experience of the *Grant* litigation, I would conversely suggest, was a valuable exercise in terms of the way in which the claim for European rights operated as, in Elizabeth Kingdom’s terminology, a “heuristic device”.¹¹⁹ The deployment of rights discourse – and most importantly its widespread publicity, at least in the United Kingdom – drew attention, not only to sexualities and relationships, but it also placed gay *women* at the forefront of a campaign. It served an important function in drawing attention at least to the possibilities and potentialities of European citizenship claims around sexuality. In a member state in which relatively few people are even aware that they are “citizens of the Union”, the *Grant* litigation made one group aware that there was a legal arena in which the language of citizenship could be articulated on their behalf.

However, as a normative matter, rights and democratic participation must operate as two strands of a political agenda. The creation of spaces of engagement traditionally has been tried – not always successfully – at the level of local, urban government.¹²⁰ But the transnational arena should not be abandoned. The fact that transnational constructs of citizenship and “belonging” have been closely linked to the market up to now, and that many in the lesbian and especially gay male communities may have little personal interest in broadening the horizon of transnational citizenship, does not mean that this should not be struggled over, so that recognition questions might be connected to issues of redistribution,¹²¹ particularly in the context of a widening European Union, and in the face of the economic disparities which inevitably will be exacerbated as a result of that expansion.

¹¹⁷ *Ibid*, at 167.

¹¹⁸ On the relationship between rights and deliberative democracy, see Benhabib, “Toward a Deliberative Model of Democratic Legitimacy” in Benhabib (ed), *Democracy and Difference: Contesting the Boundaries of the Political* (1996), p 67.

¹¹⁹ Kingdom, *supra*.

¹²⁰ See generally Cooper, *Sexing the City: Lesbian and Gay Politics within the Activist State* (1994).

¹²¹ On redistributive politics and “sexual orientation”, see Herman, “The Good, the Bad and the Smugly”, *supra*; Boyd, *supra*. On redistributive politics and sex equality, see Hervey, “Sex Equality as Substantive Justice” (1999) 62 *MLR* 614 at 622.

THE MODEL OF SOCIAL RELATIONS: STRATEGIC BEHAVIOUR, TRUST AND LOYALTY

*Joseph Savirimuthu, Senior Lecturer in Law, Lancashire Law
School.**

I. THE QUESTION

The question:

“when can shareholders bring proceedings to enforce their or
the company’s rights?”¹

may not be particularly original but it has important ramifications for both the individual and society.² Two complementary paradigms have been relied upon at common law when determining questions of standing: (i) identifying the rights of the shareholders with reference to the separate legal personality of the business medium; or (ii) articulating policy justifications permitting departure from the principle of shareholder democracy. In both instances precedent dictates that access to the courts should be the exception rather than the rule. One could reasonably base this presumption against permitting the use of the judicial forum to address intra-corporate disputes, on the judiciary’s unwillingness to become engulfed in matters raising issues where they have no relevant experience and the belief that internal governance structures were more than adequate to resolve the differences between the parties.³ This raises the obvious question, namely, what is the significance of the statutory oppression remedies and the recent proposals contained in the Law Commission Report which aims to increase transparency and accountability within business organisations by minimising procedural complexities in the current rules on standing?⁴ The specific goal of this article is to explore the critical questions of when and why the law seems particularly concerned with divergent interests in business organisations. Such questions invariably throw up a number of interesting ideas about the role of dispute mechanisms and the integrity of the internal governance framework in the business enterprise. At the heart of these ideas lie complex practical and doctrinal issues regarding the way power should be exercised,

* I am grateful to the anonymous reader for reading and commenting on an earlier version of this article. Responsibility for the final version is mine entirely.

¹ Law Commission Consultation Paper No 142, *Shareholder Remedies* (London, Stationary Office, 1996) para 1.1.

² B.L. Hay and K.E. Spier, “Settlement of Litigation” in P Newman (ed) *The New Palgrave Dictionary of Economics and the Law*, (1998) p 442.

³ B. Cheffins, *Company Law: Theory, Structure and Operation* (Oxford: Clarendon Press, 1997) pp 309-314.

⁴ Compare K. Wedderburn, “Shareholders’ Rights and the Rule in *Foss v Harbottle*”, [1957] 16 *Camb LJ* 194 and R. Drury, ‘The Relative Nature of Shareholder Rights to Enforce the Company Contract’, [1986] 45 *Camb LJ* 219 with J. Lowry, “Reconstructing shareholder actions: a response to the Law Commission’s Consultation Paper”, (1997) 18 *Company Lawyer* 247 and P. Roberts (*et al*), “Shareholder Remedies – Efficient Litigation and the Unfair Prejudice Remedy”, [1999] *JBL* 38.

made accountable and finally legitimated.⁵ This paper seeks to adopt a different approach to understanding the issues arising from the legal determination of questions of standing within the context of shareholder grievances in closely-held firms. I would like to shift the focus away from the traditional approaches to this subject which analysed questions of standing from the perspective of the relativity of the contractual rights under the common law and the curbs on opportunism under the statutory regime.⁶ Instead, I shall reorientate it on the central role of contract in the ordering of social relations in the company, the reasons for its prominence and its impact on questions of standing and more generally wealth creation.⁷ This should provide us with another source of ideas which can be utilised to enhance our understanding of the significance of the statutory developments and the recent proposals contained in the review on shareholder remedies. Before exploring these issues in greater detail, a snapshot of the historical antecedents leading to the creation of the regulatory framework of this business medium will provide us with an understanding of the underlying values and goals in the law's conception of the interactions between individual shareholders.

II. THE ECONOMIC MODEL OF WEALTH CREATION: A HISTORICAL OVERVIEW

A simple comparison between the regulatory framework governing partnerships and limited companies reveals significant differences in the legal ordering of relationships between individuals. In a partnership there is no distinction between management and its owners and in the event of any breakdown in relations, individuals can usually rely on their right to have their investment purchased.⁸ Such is not the case in respect of the company. Subject to any modifications in the constitution, the law gives particular prominence to the division of responsibilities and a hierarchy of power relations legitimated by the notion of shareholder democracy. History and economics give an insight into the reasons for this.⁹ Throughout the history of the business corporation one key challenge for the regulatory authority in the capitalist system has been to provide a framework which is both effective

⁵ See for example the essays in J. McCahery *et al.*, (eds), *Corporate Control and Accountability: Changing Structures and the Dynamics of Regulation* (Oxford, Clarendon Press, 1993); J.E. Parkinson, *Corporate Power and Responsibility: Issues in the Theory of Company Law* (Oxford, Clarendon Press, 1994) pp 3-72.

⁶ See K. Wedderburn, "Shareholders' Rights and the Rule in *Foss v Harbottle*", [1957] *Camb LJ* 194 and [1958] *Camb LJ* 93; C.A. Riley "The values behind the Law Commission's Consultation Paper", (1997) 18 *Company Lawyer* 260; W.M. Rees "Shareholder Remedies", [1997] 5 *ICCLR* 155; S. Sheikh, "Shareholder Remedies", [1997] 2 *ICCLR* 47.

⁷ H. Collins, *The Law of Contract*, (3rd ed, 1997) pp 1-45, 89-127 and 374-376; S. Wheeler and J. Shaw, *Contract Law: Cases, Materials and Commentary*, (Oxford, Clarendon Press, 1996) Chapters 2 and 3.

⁸ Partnership Act 1890, section 32.

⁹ W.R. Cornish and G. de N. Clark, *Law and Society in England 1750-1950* (London, Sweet & Maxwell, 1989) pp 1-17, 79-92; more generally A.V. Dicey, *Law and Public Opinion in England*, (2nd ed, 1963); K. Renner, *The Institutions of Private Law and their Social Functions* (London, Routledge & Kegan Paul Ltd, 1949).

and responsive to the dynamics of the market economy. That the ideology of the market should be deeply embedded in political and social spheres stems from its association with wealth creation.¹⁰ Advocates of this ideology emphasise the relationship between the pursuit of the utilitarian-liberal ideology with the advancement of an individual's self-interest on the one hand and that of the incorruptible nature of the market mechanism with the economic well-being of society on the other. The view posited is that rational individuals entering into consensual exchange transactions will lead to resources being efficiently allocated and utility maximised. It is no coincidence then that:

“the business enterprise enters the theory of the market economy, not as a social organization that takes an active role in managing its productive investments, but only as a ‘productive function’ that puts technical limits on how productive resources can be transformed into the products that consumers demand.”¹¹

It is this simplicity and apparently content-neutral character of the economic model of wealth creation in a self-regulating market which perhaps explains its appeal and uncritical acceptance in post-feudal England. The market it was thought, except in cases of ‘market failure’, by virtue of its laws and arrangements could be relied upon to coordinate the production and distribution of wealth.¹² In this period of experimentation, with no previous history of the workings of the market economy to draw upon, the elevation of the system of capitalism to a prominent role in society provoked no significant concerns from the political establishment of the day.¹³ Two observations can be made here. First, the way in which the market order and its arrangements for wealth creation serve to provide the necessary intellectual justification for limiting state intervention: interference in its ‘natural order’ and arrangements were viewed as being economically counterproductive and politically as well as socially, an unacceptable encroachment upon an individual's right to make voluntary choices.¹⁴ Second, the paradox of the instrumental role of market ideology: State patronage of the market economy was a necessary response to the constraining and limiting structures of feudal England's ability to attract

¹⁰ W. Lazonick, *Business organization and the myth of the market economy* (Cambridge Univ Press, First published 1991, paperback reprinted 1994) pp 59-67.

¹¹ *Ibid*, p 63.

¹² *Ibid*, p 108; J.H. Farrar & B. Hannigan, *Farrar's Company Law*, (4th ed, 1998) Chapters 14-17; J. Stanworth and C. Gray (ed), *Bolton 20 years on* (London, Paul Chapman Publishing, 1991) pp 1-3.

¹³ See F. Easterbrook and D. Fischel, *The Economic Structure of Corporate Law* (Cambridge, Harvard UP, 1991); cf J.A. Schumpeter, *Capitalism, Socialism and Democracy* (London, George, Allen & Unwin, 1943) and K. Polanyi, *Great Transformation* (USA, Beacon Press, 1957).

¹⁴ W.R. Cornish and G. de N. Clark, *Law and Society in England 1750-1950* (London, Sweet & Maxwell, 1989) pp 246-266; see generally B.C. Hunt, *The Development of the Business Corporation in England 1800-1867* (Cambridge, Harvard University Press, 1936).

capital.¹⁵ The emergence of the system of capitalism led to a fundamental ordering of political and social structures:

“ . . . it means no less than the running of society as an adjunct to the market. Instead of economy being embedded in social relations, social relations are embedded in the economic system. . . [f]or once the economic system is organized in separate institutions, based on specific motives and conferring special status, society must be shaped in such a manner as to allow that system to function according to its own laws.”¹⁶

What is particularly striking about the dominance of the market economy ideology during the 18th and 19th centuries, was the way it circumscribed the role of the State and legal instruments.¹⁷ The idea of a self-regulating market and the image of voluntary consensual agreements was used to support the view of the company as nothing more than a simple profit maximising vehicle.¹⁸ Hence, the focus of the policy makers was to utilise statutory legal instruments to address pragmatic concerns like establishing the most efficient way of coordinating the various factors of production and facilitating finance structures. Running in tandem with this were more general complex doctrinal concerns relating to the public/private dichotomy of the business enterprise and the legitimacy of the rules defining ownership and accountability. The emerging statutory framework reflected a preoccupation with ensuring the company could properly function in the market environment. Incorporation and the other ‘traits’ of personhood were nothing more than facilities through which the limitations of capital could be overcome.¹⁹ For example, the concept of corporate personality served as an important device to simplify and organise the various underlying realities of accountability, responsibility and ownership. Equally important was the benefit of limited liability which was designed to ensure that the company had a workable financial base from which it could maximise its wealth creating potential. Insulating risk averse shareholders (that is, equity creditors) against personal liability was but a small concession.²⁰ The priority given to the provision of a regulatory framework to facilitate the generation, maintenance and distribution of capital was designed to achieve clearly

¹⁵ K. Polanyi *Great Transformation* (USA, Beacon Press, 1957) Chapter 5 and R. Cotterrell, *The Sociology of Law* (London, Butterworths, 1992) pp 118-119.

¹⁶ *Ibid.* p 57.

¹⁷ For background information see A.B. DuBois, *The English Business Company after the Bubble Act 1720-1800* (New York, Octagon Books, 1971) and A.V. Dicey, *Law and Public Opinion in England*, (2nd ed, 1963).

¹⁸ See B. Cheffins, *Company Law: Theory, Structure and Operation* (Oxford, Clarendon Press, 1996) pp 31-46.

¹⁹ P. Ireland, “The Triumph of the Company Legal Form, 1856-1914” in J. Adams (ed) *Essays for Clive Schmitthoff* (Abingdon, Professional Books Ltd, 1983) pp 29-58; B.C. Hunt, *The Development of the Business Corporation in England 1800-1867* (Cambridge, Harvard University Press, 1936) pp 1-83; L.C.B. Gower, “A South Sea Heresy”, (1952) 68 *LQR* 214.

²⁰ B.C. Hunt, *The Development of the Business Corporation in England 1800-1867* (Cambridge, Harvard University Press, 1936) pp 116-144; see generally O.E. Williamson, *The Economic Institutions of Capitalism* (New York, 1985).

defined economic goals.²¹ The statutory provisions coupled with uncompromising judicial supervision ensured that the funds were not dissipated through self-seeking conduct which would have undermined public confidence in the State's patronage of this business medium.²² In addition to this, the regulatory framework provided "individuals with facilities for realising their wishes, by conferring legal powers upon them to create, by certain specified procedures, and subject to certain conditions, structures of rights and duties within the coercive framework of the law."²³

Individuals, it should be noted, had the choice as to whether to opt into the regulatory framework and make relevant modifications in the model of social relations provided by the law. This was not an aspect with which business seemed to be concerned as trading relations were generally conducted in the shadow of the law.²⁴ To the uninformed, however, the failure to recognise that different rules existed between a partnership and a company had important ramifications for the interaction between the participants. This was an aspect which was overseen by the judiciary.²⁵

In summary, it can be observed that the lack of direct supervision by the State after providing a framework within which interactions were conducted could be attributed to a number of reasons: the relative lack of maturity of this particular form of business enterprise and the commercial environment, the perception of entitlements to rights and interests, the belief in the limited role of the State in the market economy and the conceptualisation of the 'company' as a production mechanism.²⁶

III. THE CONTRACTUAL MODEL OF SOCIAL RELATIONS

The period between the 18th and 19th centuries can be seen as marking the transition of society from its reliance on status as a means of ordering social relations to that of contract.²⁷ One probably needs no reminder to note that the appeal of the law of contract to the wealth maximising goal of the market economy runs deep in English jurisprudence.²⁸ The idea of voluntary consensual contracting between private individuals is seen as being an efficient means of producing mutually beneficial outcomes and which in turn

²¹ A.B. DuBois, *The English Business Company after the Bubble Act 1720-1800* (New York, Octagon Books, 1971) pp 346-434.

²² See *Trevor v Whitworth* (1887) 12 App Cas 409.

²³ D. Sugarman and R. Rubin, "Introduction – Towards A New History of Law and Material Society" in G.R. Rubin and D. Sugarman (eds) *Law Economy & Society, Essays in the History of English Law 1750-1914* (Professional Books, 1984) pp 10-11.

²⁴ *Ibid.*

²⁵ See below.

²⁶ A.B. DuBois, *The English Business Company after the Bubble Act 1720-1800* (New York, Octagon Books, 1971) pp 302-307, 346-348.

²⁷ R. Cotterrell, *The Sociology of Law* (London, Butterworths, 1992) pp 118-119 and M. Cohen, "The Basis of Contract", (1933) 44 Harv L Rev. 553, 562.

²⁸ See W. Blackstone, *Commentaries on the Law of England* 4 Vols, (15th ed, 1809); P. Atiyah, *The Rise and Fall of Freedom of Contract* (Oxford, OUP, 1979); P. Hamburger, "The Development of the Nineteenth Century Consensus Theory of Contract", (1989) 7 *Law & Hist Rev* 24.

is seen as enhancing the welfare of society.²⁹ That it is viewed in largely utilitarian terms stems from its underlying characteristics: the facilitation of pareto-superior outcomes in utility maximising transactions between rational individuals, enabling the transfer of ownership of goods and services and the affinity of its formal structures with notions of individual autonomy, minimal state intrusion and the market.³⁰ The contractualisation of social relations is significant in two respects; first, it entrenches the device of contract and its underlying values and obligations to the governance of social relations and secondly, primacy is given to the obligations created therein over non-pecuniary obligations like legitimate expectations, loyalty and trust.³¹ The much heralded affinity between the model of social relations constructed by the law and the market also contains an important political message:

“. . . they provide a solution to a problem for loss of confidence in the institutionalized structures giving direction to people’s lives. The preference for markets is symptomatic of a scepticism with respect to any claim to know best how social relations should be organized, what standards they should observe and perhaps a fear that any institutionalized power of control over social relations will be abused and employed to destroy individuality.”³²

The interesting issue here is to examine the instrumental role of the contractual ordering of social relations in the company. The first point that one can make is that the perception of obligations between shareholders being consensual and voluntary is turned on its head by the imposition of a statutory contract which dictates relations within the company.³³ Section 14 of the Companies Act 1985 provides that membership subjects all shareholders to a binding contract. The ordering of relations through this device is significant in a number of respects.³⁴ First, the obligations and rights between the parties are determined by reference to the company’s

²⁹ See Jensen and Meckling, “Theory of the Firm: Managerial Behaviour: Agency Costs, and Ownership Structure”, (1976) 3 *J Fin Econ* 305; H. Collins, *The Law of Contract*, 3rd ed (London, Butterworths, 1997) pp 3-10; F.S. McChesney “Legal Change and Small Business Law”, (1997) 1 *Journal of Small and Emerging Business Law* 1 at 5.

³⁰ R Coase, “The Problem of Social Cost”, (1960) 3 *J Law & Economics* 1; Calabresi and Melamed, “Property Rules, Liability Rules, and Inalienability: One view of the Cathedral”, (1972) 85 *Harv L Rev* 1089; but see V. Brudney, “Corporate Governance, Agency Costs and the Rhetoric of Contract”, (1985) 85 *Colum L Rev* 1085, De Mott, “Beyond Metaphor: An Analysis of Fiduciary Obligation”, [1988] *Duke LJ* 879 and V. Brudney, “Contract and Fiduciary Duty in Corporate Law”, (1997) 38 *Boston College L Rev* 595.

³¹ H. Collins, *The Law of Contract*, (3rd ed, 1997) p 89. This does not however preclude the deployment of fiduciary principles. Such principles could be broadly construed as safeguarding the economic interests of the ‘*persona ficta*’. Breaches of fiduciaries could only be enforced by the ‘company’ in general meeting.

³² H. Collins, *The Law of Contract*, (3rd ed, 1997) p 90.

³³ Individuals incorporate with little understanding of the terms of the contract and the significance of the articles of association. The early jurisprudence of the contractual effect of the articles of association reflects a feature of the common law – providing solutions to problems as they accrued.

³⁴ *Ibid* pp 102-104.

memorandum and articles of association but in itself do not give rise to legally enforceable remedies. The law to this end is preoccupied with giving effect to the 'bargain' deemed to be entrenched in the company's constitution and hence as a consequence attaches little importance to any 'legitimate expectations' that may have been created by virtue of the parties ongoing relationship. Second, the imposition of the contract device 'depersonalises' relationships within the business venture. For instance, the model of social relations defines individuals as 'director' or 'shareholder' and characterises their respective rights and obligations accordingly. This has the effect of rendering the personal circumstances, size of investments or expectations virtually superfluous to any determination of rights or remedies. For example, the 'share' entitles its owner only to 'a bundle of rights', but its precise scope and nature is left ambiguous.³⁵ The interests of the shareholder are viewed primarily in quantifiable terms with little or no accommodation of expectations like trust, duty or loyalty. Amidst this ordering of social relations the '*persona ficta*' and not the shareholder is treated as the primary beneficiary of many of the rights, duties and obligations. Finally, the contractualisation of social relations leads to a hierarchy of power relations supplemented by rules and structures which are elevated to a constitutional status. It is worth reminding ourselves of the extent of this 'distortive' effect on social relations by drawing a comparison with the ordering of relationships in a partnership. In this business medium even though the relationship is governed by contract it does not lead to the creation of a unilateral system of rule making or a hierarchy of power relations.³⁶ All individuals are partners and subject to contrary provisions, have rights and obligations which are mutually enforceable and the social cohesion is provided by the overarching obligations of good faith, trust and loyalty. This is in clear contrast with the accompanying state of affairs if such individuals were to incorporate their business. For instance, shareholders in general meeting are vested with the responsibility of exercising stewardship over the moral hazard and principal-agency problems that emerge from the allocation of management functions to the board of directors.³⁷ This model of social ordering is adhered to even though no distinction between ownership and management may exist. The restrictive interpretation of the statutory contract and the limited avenue for enforcing rights is usually justified on efficiency grounds, citing the economic consequences of multiple lawsuits, 'free-rider' problems, the avoidance of disruption of the business and the detrimental impact of litigation on relationships.³⁸ DuBois affirms the principle of shareholder democracy on the grounds that:

³⁵ P. Ireland, "Company Law and The Myth of the Shareholder Ownership" (1999) 62 *MLR* 32.

³⁶ G. Morse, *Partnership Law*, (3rd ed, 1986) pp 108–142.

³⁷ See *Automatic Self-Cleansing Filter Syndicate Co Ltd v Cunningham* [1906] 2 Ch 34; *Hickman v Kent* [1915] 1 Ch 881; *Quin & Axtens Ltd v Salmon* [1909] 1 Ch 311; *Beattie v E & F Beattie Ltd* [1938] Ch 708; Y. Kotowitz, "Moral Hazard", in J. Eatwell, Milgate and P Newman, (eds) *The New Palgrave: A Dictionary of Economics* vol 3 (1987) pp 549-551; *Boardman v Phipps* [1967] 2 AC 46; E.J. Weinrib, "The Fiduciary Obligation", (1975) 24 *Univ Toronto LR* 1.

³⁸ L.S. Sealy, "The Enforcement of Partnership Agreements, Articles of Association and Shareholder Agreements" in P. Finn (ed) *Equity and Commercial Relationships* (Sydney, Law Book Co, 1987) p 89; R. Drury, 'The Relative Nature

“. . . rule by the majority was the essence of the organization. In charters, in articles of association, in acts of incorporation, there were to be found numerous provisions which sought to guarantee the advantages of speed and decision that majority control made possible.”³⁹

It should be noted that in the general scheme of things the contractualisation of social relations does work – otherwise no one will wish to be a minority shareholder.⁴⁰ This does not detract in any way from our highlighting the bias in the contractual model towards economic values and goals and its implications for the ordering of social relations particularly when individuals decide to act strategically.

Before proceeding to identify the reasons motivating decisions to act strategically, it is important to recognise the significance of traditional judicial perceptions of their role in the processing of disputes between shareholders. Judicial sympathies with the integrity of the contractual ordering of social relations is reaffirmed by the erection of procedural barriers to inhibit litigation. This is commonly known as the rule in *Foss v Harbottle*.⁴¹ The minority shareholders in this company were concerned about what they perceived as a clear abuse by its directors of their duties resulting from them making a profit on a sale of property to the company. The response of Wigram V-C was clear and uncompromising. It was held that in respect of any improprieties or mismanagement within the company, individual shareholders were not legally recognised as beneficiaries entitled to redress on the basis of the privity doctrine. This, the court stressed, was consistent with the bargain reached under the statutory contract. Following from this, the rule states that only the ‘company’ through its proper constitutional body is entitled to either pursue relief through the courts or ratify the impropriety concerned. As the Court of Appeal in *Prudential Assurance Co Ltd v Newman Industries (No2)* noted “. . . [the] rule is not merely a tiresome procedural obstacle placed in the path of a shareholder by a legalistic judiciary. The rule is the consequence of the fact that a corporation is a separate legal entity. . .”⁴².

The exceptions to the presumption against legal intervention, in particular the ‘fraud on the minority’ exception, reflects judicial acknowledgment of the limits to the autonomy of the contractual model of social relations. A clear exposition of the limits to the private ordering of individual relations can be found in the judgment of Jenkins LJ in *Edwards v Halliwell*.⁴³ In essence, judicial intervention was deemed to be justified in cases where it could be shown that the conduct of the directors / majority shareholders pointed to a

of a Shareholder's Right to Enforce the Company Contract’, [1986] 45 *Camb LJ* 219.

³⁹ A.B. DuBois, *The English Business Company after the Bubble Act 1720-1800* (New York, Octagon Books, 1971) pp 302-303.

⁴⁰ T. O’Neil, “Self-Interest and Concern for Others in the owner-managed firm: A Suggested Approach to Dissolution and Fiduciary Obligation in Close Corporations” [1992] 22 *Seton Hall L Rev* 646, 649.

⁴¹ 2 Hare 461; also *Mozley v Alston* (1847) 1 Ph 790.

⁴² [1982] Ch 204, 223.

⁴³ [1952] 2 All ER 1064, 1067. See also *Prudential Assurance Co Ltd v Newman Industries (No2)* [1982] Ch 204.

clear departure from accepted social norms and values.⁴⁴ For example, a shareholder could pursue a personal action if he could point to the infringement of a legally recognised personal right, as was the case in *Pender v Lushington*.⁴⁵ Jessel MR in this case was of the view that a denial of a right to vote was an individual right which could not be encroached upon by the majority in the company. It should however be stressed that as the scope and nature of the rights which were enforceable under the statutory contract were narrowly defined, recourse to the courts through this avenue was illusory.⁴⁶ In cases where there had been a breach of duty on the part of the directors, resulting in loss to the company's assets, it was made clear that with the exception of fraud, mere negligence in itself did not give rise to any personal remedy for the shareholder.⁴⁷ Judicial intervention at best can be described as being selective; access to the courts seems broadly to be permitted if failure on the part of the 'company' to seek redress is viewed as being tantamount to a failure to exercise voting powers responsibly or where the benefits accruing to company officers is an equitable fraud on the company.⁴⁸ Any suggestion that these interventions might provide the momentum for greater judicial activism was short circuited by the raising of additional formalist barriers like obtaining independent ratification and judicial approval in funding litigation costs.⁴⁹ For example, in *Smith v Croft (No 2)* the mere fact of there being evidence of 'fraud' was regarded by the court as being insufficient to permit legal intervention.⁵⁰ The view of Knox J was that:

“[u]ltimately the question which has to be answered in order to determine whether the rule in *Foss v Harbottle* applies to prevent a minority shareholder seeking relief as plaintiff for the benefit of the company is 'Is the plaintiff being improperly prevented from bringing these proceedings on behalf of the company?' If it is an expression of the corporate will of the company by an appropriate independent organ that is preventing the plaintiff from prosecuting the action he is not

⁴⁴ See *Cook v Deeks* [1916] 1 AC 554; *Daniels v Daniels* [1978] Ch 406.

⁴⁵ (1877) 6 Ch D 70.

⁴⁶ *Wood v Odessa Waterworks Co* (1889) 42 Ch D 636, *Pender v Lushington* (1877) 6 Ch D 70; also see *Prudential Assurance Co v Newman Industries Ltd (No2)* [1982] 1 All ER 354; *Stein v Blake and others (No2)* [1998] 1 BCLC 573. However, in the ruling by the Court of Appeal in *Barings plc (in admin) v Coopers & Lybrand (a firm)* [1997] 1 BCLC 427 it was observed that individuals could have a legally recognisable interest in relation to the diminution in the value of the shareholding if a duty of care was shown to be present.

⁴⁷ *Pavlides v Jensen* [1956] 2 All ER 518

⁴⁸ *Daniels v Daniels* [1978] Ch 406; *Estmanco (Kilner House) v GLC* [1982] 1 WLR 2.

⁴⁹ See *Wallersteiner v Moir (No2)* [1955] QB 373; *Smith v Croft (No2)* [1998] Ch 114; L.S. Sealy, "Problems of Standing, Pleading and Proof in Corporate Litigation" in B. Pettet (ed) *Company Law in Change* CLP (London, Stevens, 1997); *Prudential Assurance v Newman* [1982] Ch 204; see A. Boyle, "Judicial Policy and the Derivative Action" in *Developments in European Company Law* Vol 1 Barry Rider and Mads Andenas (ed) (London: Kluwer, 1996) p111. See also the ruling in *Barrett v Duckett* [1995] 1 BCLC 73 which increases the threshold for reliance on the 'fraud on the minority' exception.

⁵⁰ [1988] Ch 114.

improperly but properly prevented and so the answer to the question is 'No'."⁵¹

One might conclude from the two extracts cited above that the law is “. . . objective and neutral, not a matter of politics or preference, but a settled body of rules and principles, legitimated by tradition and routine observance, and applied impartially and fairly to all citizens.”⁵²

It has already been shown however, that this is not necessarily the case given the close relationship between the regulatory framework and the functioning of the market.⁵³ For instance, the formalist approach to the processing of disputes assumes that interactions exist in a purely rational construct. This could be construed as evidence of judicial perception of the market environment rather than litigation as being the appropriate mechanism for maintaining discipline and order. There is however one serious issue that has received little attention. The marginal role of both the State and the courts in the private ordering of individual affairs has traditionally been perceived as being necessary, so as not to undermine the arrangements of the market order. The problem with this view is that it treats the market environment as being capable of maintaining stable as well as rational conditions within the company. To this end the contractual model of social relations assumes an important instrumental role. Is it entirely accurate to say that this model with its underlying economic values and norms is responsible for the maintenance of stability amongst the individuals? This question can be approached in another way. What if it is shown that the emerging institutional role reflects an increasing awareness of the failure of this basic model to maintain the 'rational conditions'? What does this indicate in terms of the general policy and the modern role of legal instruments? It is these and related questions that the remainder of the paper seeks to address.

IV. INSTITUTIONALISM AND THE PROBLEM OF SOCIAL RELATIONS

It will be apparent from the previous discussion that there is an inextricable link between the primacy given to contractual values and norms on the one hand and judicial perceptions of their role in processing disputes on the other. These seemingly logical and necessary rules contain:

“. . . two principal dangers which may be provoked by experimentation through contracts. In the first place, law of contract analyses and evaluates social relations from within its confines by reference to criteria which may lead to an undesirable distortion of the meaning of the social relation involved. Secondly, the seductive attraction of avoidance of explicit distributive decisions by delegating outcomes to the

⁵¹ *Ibid* 185.

⁵² H. Collins, *The Law of Contract* (3rd ed, 1997) p 7; I.R. Macneil, *The New Social Contract, An Inquiry into Modern Contractual Relations* (Yale Univ Press, 1980) pp 93-94.

⁵³ M.J. Whincop, "Overcoming Corporate Law: Instrumentalism, Pragmatism and the Separate Legal Entity Concept" (1997) 15 *Comp and Securities LJ* 411, 411-413.

market mechanism can pose a threat to important principles of distributive justice.”⁵⁴

One archetypal complaint in a closely-held firm – denial of legitimate expectations – can be used to illuminate these concerns.⁵⁵ A business run by two individuals is incorporated. Both parties become directors and shareholders in the company. It is later agreed that a friend of one of the directors should be brought into the business. This individual is appointed as director and acquires shares in the company. The founders of the business have a falling out. The two friends exercise their rights under the Companies Act 1985, section 303, removing the co-founder as director. The operation of the principle of shareholder democracy and the rule in *Foss v Harbottle* means that the issues arising here will be determined by a literalist construction of the ‘bargain’.

This broadbrush approach to social relations under the contract model attaches little significance to legitimate expectations or the fact that when interests within businesses conflict parties should be expected to reach a compromise. The culture of profit maximisation seems ill-equipped to accommodate the notion that the company is also a social organisation. If one accepts that contract as a social ordering mechanism suffers from this and other failings we need to assess the particular significance of the modern role of legal instruments, notably section 459 of the Companies Act 1985 and to a lesser extent section 122(1)(g) of the Insolvency Act 1986 in this context. In contrast to the common law’s stance on upholding the primacy of the governance model of social relations, these provisions at first blush can be viewed as providing an important avenue for shareholders to subject the decision-making process within the model of social relations to judicial scrutiny.⁵⁶ Minority shareholders now have ready access to the courts in situations where divergences in interests lead to majority shareholders using their voting powers to either enforce a literalist construction of the statutory contract or to depart from previous understandings. The importance of the institutional dimension to the private ordering of individual affairs lies in the way it questions traditional conceptions of the ‘company’, ‘interests’ and the approach to wealth creation.⁵⁷ In *Re a Company (No 00477 of 1986)* Hoffmann J (as he then was) stressed that the unfair prejudice provision enabled the court to take into account wider equitable considerations when scrutinising relationships within the company.⁵⁸ In *Re Saul D Harrison and Sons plc* it was noted that:

“ . . . [conduct] can also be unfair without being unlawful. In a commercial context, this may at first seem surprising. How can it be unfair to act in accordance with what the parties have

⁵⁴ H. Collins, *The Law of Contract* (3rd ed, 1997) p 102.

⁵⁵ E. Boros, *Minority Shareholders’ Remedies* (Oxford, Clarendon Press, 1995) pp 5-8.

⁵⁶ See the Cohen Report Cmd 6659 which resulted in the passing of section 210 Companies Act 1948.

⁵⁷ *Scottish Co-operative Wholesale Society Ltd v Meyer* [1958] 3 All ER 66; *Re HR Harmer* [1959] 1 WLR 62. Also E. Boros, *Minority Shareholder Remedies* (Oxford: Clarendon Press, 1995) pp135-155; see *Re Saul D Harrison* [1995] 1 BCLC 14 and *Re Elgindata Ltd* [1991] BCLC 959.

⁵⁸ 2 BCC 99; see also *Ebrahimi v Westbourne Galleries Ltd* [1973] AC 360.

agreed? As a general rule, it is not. But there are cases in which the letter of the articles does not fully reflect the understandings upon which the shareholders are associated. . . Thus the personal relationship between a shareholder and those who control the company may entitle him to say that it would in certain circumstances be unfair for them to exercise a power conferred by the articles upon the board or the company in general meeting. . . [a legitimate expectation] often arises out of a fundamental understanding between the shareholders which formed the basis of their association but was not put into contract form.”⁵⁹

One detects from this passage not only a fundamental shift away from the previously uncritical attitude towards the primacy of contract ideology but a heightened awareness of the organisational dimension of the ‘family’ or ‘lifestyle’ company and the value of social norms in the wealth creation process.⁶⁰ It is important at this juncture to make clear for present purposes what we mean when we say a breach of ‘trust’ or a departure from ‘social norms’ gives rise to a remedy under the statutory regime. For example, a dictionary definition of ‘trust’ is “[f]aith, or confidence in the loyalty, strength, veracity. . . of a person.”⁶¹

In company law however, it has been made clear that not every reliance, divergence in interest, allegations of unfairness or prejudicial conduct will give rise to a legal remedy.⁶² As the Court of Appeal recently indicated:

“In deciding what is fair or unfair for the purposes of s.459, it is important to have in mind that fairness is being used in the context of a commercial relationship. The articles of association are just what their name implies: the contractual terms which govern the relationships of the shareholder with the company and each other. Since keeping promises and honouring agreements is probably the most important element of commercial fairness the starting point in any case under s.459 will be to ask whether the conduct of which the shareholder complains was in accordance with the articles of association. A finding that conduct was not in accordance with the articles does not necessarily mean that it was unfair still less that the court will exercise its discretion to grant relief.”⁶³

This is an important judicial acknowledgment that divergences of interests invariably arise in business relations and that the law should not be utilised to

⁵⁹ [1994] BCC 475, 489-90.

⁶⁰ See *Re Postgate & Denby (Agencies) Ltd* [1987] BCLC 8; *Re Cumuna Ltd* [1986] BCLC 430; *Re a Company* [1986] BCLC 376.

⁶¹ *The New Shorter Oxford English Dictionary* Volume 2, L Brown (ed) (Oxford, Clarendon Press, 1993) p 3411.

⁶² *Re a Company* [1986] BCLC 362; *Re London School of Electronics* [1985] 3 WLR 474; E. Boros, *Minority Shareholder Remedies* (Oxford: Clarendon Press, 1995) pp 128-152.

⁶³ *Re Saul D Harrison* [1994] BCC 475, 488-489.

reduce business relations to legal propositions.⁶⁴ This much is evident from the observations of the House of Lords in *Ebrahimi v Westbourne Galleries*.⁶⁵ The Law Lords in subjecting the exercise of legal rights to equitable considerations stressed that this could only be justified where the act or omission was tantamount to a departure from certain fundamental understandings contrary to good faith.⁶⁶ More recently, Lord Hoffmann noted in *Re a Company* that the mere fact that trust and confidence between the parties had broken down did not invariably give rise to a legal remedy.⁶⁷ Individuals in business must accept the risk that at times divergences of interests may place them at a disadvantage. Where the petitioner is successful, in contrast to the position at common law, the court has extensive powers which include the right to institute a derivative suit, order a 'buy-out' or regulate the affairs of the company.⁶⁸

It will be observed that the proactive institutional role has led to a shift in focus away from the notion of contract as a primary governance model. This is significant in a number of respects – first, its recognition that the individual shareholder is distinct from the company and capable of having expectations and interests which extend beyond the economic rights associated with share ownership.⁶⁹ Second, it permits claims by individual shareholders even though technically the company has a similar entitlement.⁷⁰ Third, it provides a solution to the adverse consequences accompanying the imposition of a hierarchy of power relations in the company.⁷¹ Individual shareholders in closely-held firms can obtain a remedy based on an objective standard of fairness and hence not dependent on proof of *mala fides*.⁷² For completeness, it should be pointed out that the limitations of the model of social relations is not particularly problematic in the case of minority shareholders in companies with public listings. Values like trust and loyalty seem not to fit in readily in an environment where there is an automatic exit remedy and shareholdings are generally viewed as investment opportunities. This much seems to be implied in the observation by Jonathan Parker J in *Re Astec (BSR) plc*:

“... the concept of ‘legitimate expectations’... can have no place in the context of public listed companies. Moreover, its introduction in that context would, as it seems to me, in all probability prove to be a recipe for chaos. If the market in a

⁶⁴ D. Sugarman and R. Rubin, “Introduction – Towards A New History of Law and Material Society” in G.R. Rubin and D. Sugarman (eds) *Law Economy & Society, Essays in the History of English Law 1750-1914* (Professional Books, 1984).

⁶⁵ [1973] AC 360; B. Rider, “Partnership Law and the Impact on Domestic Companies” [1979] *CLJ* 148.

⁶⁶ [1973] AC 360, 379.

⁶⁷ [1999] 2 All ER 961, 973.

⁶⁸ See the remedies available under Company Act 1985, section 461; *Re Postgate & Denby (Agencies) Ltd* [1987] BCLC 8; *Re London School of Electronics* [1985] 3 WLR 474.

⁶⁹ *Re a Company No 002567 of 1982* [1983] 1 WLR 927.

⁷⁰ See *Prudential Assurance Co Ltd v Newman Industries Ltd (No2)* [1982] 1 All ER 354.

⁷¹ *Re a Company* [1986] BCLC 362.

⁷² *Re R.A. Noble & Sons (Clothing) Ltd* [1983] BCLC 273.

company's shares is to have any credibility, members of the public dealing in that market must it seems to be entitled to proceed on the footing that the constitution of the company is as appears in the company's public documents, unaffected by any extraneous equitable considerations and constraints."⁷³

V. THE LAW COMMISSION REVIEW: A BRIEF COMMENT

Any consideration of the model of social relations will not be complete without some reference to the Law Commission review with a view to highlighting the broader context within which it operates and its implications for approaches to wealth creation.⁷⁴ An appreciation of this important development is necessary for three reasons, first: it draws our attention to the complex commercial environment within which regulations operate; second, to the increasing role of the State and finally, it reminds us that the increasing recourse to statutory legal instruments culminating in the recent review of 'Shareholder Remedies' is not the result of chance or coincidence. A number of factors can be relied upon to support the above assertions. First, the willingness of the judiciary to make the unfair prejudice remedy effective has had adverse implications. Disputes involving minority shareholders invariably lead to a thorough investigation of the history of the business relationship including the purpose of the venture and the conduct of the parties. This has a clearly detrimental effect on the business for obvious reasons.⁷⁵ As a consequence parties have had to take into account the financial and economic consequences of instituting or defending legal proceedings. In *Re Elgindata* the order sought was a purchase of shares whose original value was £40,000. The proceedings went on for 43 days giving rise to costs of £320,000 and an order to purchase the shares at £24,600 was made.⁷⁶ The cost implications to individuals must also take into account the increasing burdens on the resources of the legal system. Second, there is also the wider commercial environment to consider. Increased reliance on public funds by institutional investors when investing in listed companies has raised another spectre – public scandals.⁷⁷ Institutional pension and insurance fundholders now own close to 50% of the equity, in companies listed on the stock exchange. It has not escaped the reformers notice of the unsuitability of the unfair prejudice provisions and the problems in applying for a *Wallersteiner* order where the petitioner was an institutional shareholder.⁷⁸ Third, the reform agenda owes its impetus to the publication

⁷³ (1999) BCC 59, 87; see also *Re Blue Arrow plc* (1987) 3 BCC 618, 623 per Vinelott J, *Leeds United Holdings plc* [1997] BCC 137. See for example the valuable role of derivative actions in this respect: A Boyle, 'Judicial Policy and the Derivative Action' in *Developments in European Company Law Vol 1* Barry Rider and Mads Andenas (eds) (Kluwer, 1996) 111. Also for a comparative perspective see: GP Stapledon 'Use of the Oppression Provision in Listed Companies in Australia and the United Kingdom', (1993) 67 *Aust LJ* 575.

⁷⁴ Law Commission Consultation Paper No 142, *Shareholder Remedies* (London: HMSO, 1996).

⁷⁵ *Re Elgindata Ltd* [1991] BCLC 959.

⁷⁶ See *Re Macro (Ipswich) Ltd* [1994] 2 BCLC 354.

⁷⁷ See however *Re Astec (BSR) plc* [1998] 2 BCLC 556.

⁷⁸ D. Faber, "Reform of Shareholder Remedies" in *Developments in European Company Law Vol 1* Barry Rider and Mads Andenas (eds) (Kluwer, 1996) p 119,

of the Report on Access to Justice.⁷⁹ Lord Woolf's proposals which have been extended to the field of company law could be viewed as the 'visible hand' designed to minimise the strains on the public purse as well as improving access to justice.⁸⁰ The significant point here is that in answering the question regarding standing, we can see the way the underlying values and goals of the market mechanism have permeated discussions on the function and role of 'private' and 'public' law in the civil justice system.⁸¹ As mentioned previously, shareholders' decisions to litigate not only impose costs on private individuals and the company but on society in terms of the costs of funding the legal system.⁸² Fourth, there is also a 'political' dimension in the move towards greater 'institutional' involvement in the private ordering of individual relations.⁸³ The UK's membership in the European Union, for example, has seen greater moves towards harmonisation to facilitate the functioning of the single market.⁸⁴ The Labour government has shown a commitment to providing a regulatory framework "... which is up-to-date, competitive and designed for the next century, a framework which facilitates enterprise and promotes transparency and fair dealing."⁸⁵

at 123; see A. Boyle, "Judicial Policy and the Derivative Action" in *Developments in European Company Law Vol 1* Barry Rider and Mads Andenas (eds) (Kluwer, 1996) p 111, at 117.

⁷⁹ Lord Woolf, *Access to Justice, Final Report to the Lord Chancellor on the Civil Justice System in England and Wales* (London, HMSO, 1996).

⁸⁰ See trenchant criticism of Lord Woolf's proposals by M. Zander, "Access to Justice – Towards the 21st Century" in *Law, Society, and Economy* R. Rawlings (ed) (Oxford, Clarendon Press, 1997) 339. General discussion of these proposals can be found in A. Zuckerman and R. Cranston (eds) *Reform of Civil Procedure* in (Oxford: Clarendon 1995); P Fenn (*et al*) "Long-term Contracts in the Market for Health Care" in *Contract and Economic Organisation*, D. Campbell and P. Vincent-Jones (eds) (Dartmouth, 1996) 143. Law Commission Consultation Paper No 142, *Shareholder Remedies* (London: The Stationery Office, 1996); J. Eaglesham and J. Mason, "Legal system poised for 'big bang' overhaul from today", *Financial Times* April 26 1999, 10; J.A. Schumpeter, *Capitalism, Socialism and Democracy* (London, George, Allen & Unwin, 1943) p 157.

⁸¹ See for example the collection of papers in P. Birks (ed), *Wrongs and Remedies in the Twenty-First Century* (Oxford, Clarendon Press, 1996).

⁸² N. Rickman, "The Economics of Cost-shifting Rules" in A. Zuckerman and R. Cranston, *Reform of Civil Procedure* (Oxford, Clarendon, 1995) p 327.

⁸³ J. Gray, *False Dawn* (London, Granta Publications, 1998); W. Hutton, *The State We're In* (London, Vintage 1996); G. Soros, *The Crisis of Global Capitalism* (London, Little, Brown and Co 1998).

⁸⁴ Buxbaum and Hopt, *Legal Harmonisation and the Business Enterprise* (European Univ Institute, 1988); F. Wooldridge, *Company Law in the UK and the EC* (London, Athlone Press, 1991); D. Wyatt (*et al*) *European Community Law* (London: Sweet & Maxwell 1993); DTI, *Business In Europe, The Single Market, Company Law Harmonisation* (London: 1993). See also J. Kelly, 'Brussels accounting plans worry top companies' *Financial Times* 20/21 Feb 1999 at p6.

⁸⁵ DTI, *Modern Company Law For a Competitive Economy*, (London, HMSO, 1998); Board of Trade, *Report of the Company Law Committee* (London, HMSO, Cmnd 1749); *Report of the Committee of Inquiry on Small Firms: Small Firms* (London, HMSO, Cmnd 4811, 1971); J. Stanworth and C. Gray (eds), *Bolton 20 years on* (London, Paul Chapman Publishing, 1991); J. Martinson, "Review of company law revives 'stakeholder' issue", *Financial Times* February 25 1999 1; J. Kelly and J. Martinson, "Epic act to pull company law out of Victorian era",

The result of the review will see the rule in *Foss v Harbottle* being abrogated and in its place a new statutory derivative action, the present legal burden in petitions under section 459 being reversed and greater emphasis being placed on active case management and encouraging parties to utilise 'self-help' remedies. A survey of the existing literature shows that the response to the proposals has been far from favourable.⁸⁶ Criticisms have already been directed towards the Commission's failure to accurately 'conceptualise' the nature of the problems, its broadbrush approach to intra-corporate grievances, the surprising omission of developments in jurisdictions within the European Union and the surprising segregation of the enforcement machinery and directors' duties. It is beyond the scope of this article to address the issues arising therefrom. Given the nature of the proposed reforms, it is worth exploring one aspect of the emerging role of legal instruments and the recent developments in the model of social relations.

VI. TRUST, DUTY AND LOYALTY AS 'CAPITAL'

The theory of economic modelling and its influence in the basic model of social relations constructed by the law makes two fundamental assumptions: first, that rational behaviour is solely concerned with the productive/profit maximising goal of the company and second, the central role of the contractual model in maintaining 'rational conditions' within the company.⁸⁷ Such assumptions reflect the ". . . failure. . . in the new institutional economics to recognize the importance of concrete personal relations and networks of relations. . . in generating trust, in establishing expectations, and in creating and enforcing norms."⁸⁸ This illustrates a tendency to treat such values and norms as being nothing more than "a sort of ever-ready lubricant

Financial Times 26 February 199; The Report of the Committee on the Financial Aspects of Corporate Governance (London, 1992); Director's Remuneration: The Report of a Study Group chaired by Sir Richard Greenbury (London, 1992); The Law Society, *Memorandum by the Society's Standing Committee on Company Law: A New Form of Incorporation for Small Firms* (The Law Society, No 79, 1981); Forum of Private Business, *A Report into Business Legal Structures* (London, 1991) DTI, *Company Law Review: The Law Applicable to Private Companies* (London: DTI, URN 94/529, 1994).

⁸⁶ D. Sugarman, "Reconceptualising company law: reflections on the Law Commission's consultation paper on shareholder remedies: Part 1" (1997) 18 *Company Lawyer* 226; "Part 2" (1997) 18 *Company Lawyer* 274; J. Lowry "Reconstructing shareholder actions: a response to the Law Commission's Consultation Paper" (1997) 18 *Company Lawyer* 247; A.J. Boyle "The new derivative action" (1997) 18 *Company Lawyer* 256; C.A.Riley "The values behind the Law Commission's consultation paper" (1997) 18 *Company Lawyer* 260; L.J. Moran "Missing links and missed opportunities" (1997) 18 264; W.M. Rees "Shareholder Remedies" [1997] 5 *ICCLR* 155; S. Sheikh, "Shareholder Remedies" [1997] 2 *ICCLR* 47; P. Roberts (*et al*), "Shareholder Remedies – Efficient Litigation and the Unfair Prejudice Remedy" [1999] *JBL* 38; *cf* D. Capper, "Keeping Woolf From the Door – The Reform of Civil Procedure in Northern Ireland" (1999) 50 *NILQ* 434.

⁸⁷ B. Chapman, "Trust, Economic Rationality, and the Corporate Fiduciary Obligation" (1993) 43 *Univ Toronto LJ* 547,579.

⁸⁸ J. Coleman, *Foundations of Social Theory*, (Cambridge, Harvard UP, 1990) p 302.

that permits voluntary participation in production and exchange.”⁸⁹ However focussing on the company as a ‘social organisation’ enables us to expose the limitations of these assumptions. For instance, it is accepted that relationships between individuals in companies are conducted amicably. If conflicts of interests emerge, it is reasonable to assume that ‘rational’ individuals will not be preoccupied with profit maximising goals. Rather, every effort will be made to reach a consensus or compromise through recourse to established social norms and values. Interestingly, where such norms and values cease to have any stabilising influence, it is because of the fact that the relationship has already run its course. In such instances, majority shareholders resort to the contractual model of social ordering.⁹⁰ Viewing business as a game will help illuminate the points made above.⁹¹ Shareholders in closely-held firms can be seen as players engaged in strategic decision-making in an environment of interdependence. Each individual during the course of the business relationship regularly evaluates his goals, the conduct and reactions of the other parties and the strategies which maximises his payoff.⁹² The decisions each individual makes will be influenced by his perception and understanding of what the other participants are likely to decide and the options open to them. As each player is ‘rational’ he will opt for those strategies which offer him the best payoff.⁹³ Where relations between the players are harmonious, the strategic interactions are unlikely to give rise to coordination problems resulting in litigation. The ‘common’ or ‘group interest’ it could be assumed will be maintained, so long as the players perceive norms and values like cooperation, trust and loyalty as being crucial to the continuance of the business relationship. This perhaps explains in part, the valuable role of social norms and values in addressing conflict situations in the business relationship. It is to be noted, however, that the ordering of social relations in these ‘rational conditions’ is being achieved by non-economic values and processes rather than through the contractual model of social relations.⁹⁴ It is a trite point that social relations are never *ceteris paribus*. Competing goals, expectations or changed circumstances in situations where there has been an irretrievable breakdown in relations or where the pursuit of self-interest is viewed as being more important than the common interest, results in the

⁸⁹ P. Dasgupta, “Trust as a Commodity” in *Trust: Making and Breaking Cooperative Relations* Diego Gambetta (ed) (Basil Blackwell, 1990) p 49.

⁹⁰ B. Cheffins, *Company Law: Theory, Structure, and Operation* (Oxford, Clarendon Press, 1997) Ch 6; J. McMillan, *Games, Strategies and Managers* (Oxford, 1992) p 25; A. Katz, “The Strategic Structure of Offer and Acceptance: Game Theory and the Law of Contract Formation”, (1990) 89 *Michigan LR* 216.

⁹¹ J. McMillan, *Games, Strategies and Managers* (Oxford, 1992) pp 3-6.

⁹² *Ibid*, pp 25-27; R. Cooter, “The Cost of Coase”, (1982) 11 *Journal of Legal Studies* 1, 1-29.

⁹³ O. Young, ‘Introduction’, Oran R. Young (ed) *Bargaining, Formal Theories of Negotiation* (Chicago, Univ of Illinois Press, 1975) 22-24; J. McMillan, *Games, Strategies and Managers* (Oxford, 1992) pp 3-6.

⁹⁴ I. MacNeil, “Economic Analysis of Contractual Relations: Its Shortfalls and the Need for a ‘Rich Classificatory Apparatus’ ”, (1981) *NW Univ L Rev* 75, 1047-1048.

parties reverting to the contractual model.⁹⁵ The primacy accorded to majority rule creates a zero-sum game whereby one party always wins and the other loses. As Collins observes, where there has been a divergence of interest, the contractual model assumes strategic proportions as it:

“permits each group to pursue its interests without any concern for the other parties who are bound to the firm. Once the articles of association become regarded as simply another contract, this time between shareholders, then this contractual framework destroys any sense that the corporate entity comprises an organization, which has to balance and protect the interests of various constituencies or members.”⁹⁶

There are two ways of viewing the limited intervention by the State in the provision of minority oppression remedies.⁹⁷ One is that the institution is being relied upon to ensure as far as possible that ‘rational conditions’ are maintained; the other, as a remedy available in respect of departures from accepted social norms and values. That the model of social relations suffers from a ‘trust deficit’ has to some extent been understood by shareholders in closely-held firms, in view of the increasing recourse to legal advice and shareholder agreements.⁹⁸ The emergence of statutory legal instruments reflects an important role being assumed by the law. It is indisputable that legal instruments which increase accountability and transparency in the decision-making process in companies, are capable of creating conditions where norms and values like ‘trust’, ‘duty’ and ‘loyalty’ can thrive. What significance can we attach to these institutional developments in the wider policy of wealth creation? It will be observed that under the conventional approach to wealth creation, *capital* contributions were seen as its mainstay. A symbiotic relationship seems to have been accepted almost uncritically between the orthodox notion of capital, with its emphasis on brick, motor, plant, buildings and money and the conceptualisation of the ‘company’ and the roles of legal instruments and structures in the ordering and governance of social relations.⁹⁹ Ingrained in this process of wealth creation is a closed system of thought involving “. . . mindsets. . . stuck in an old model, in

⁹⁵ F. Fukuyama, *Trust: The Social Virtues and the Creation of Prosperity* (London, Hamish Hamilton, 1995) pp 3-4; B Chapman, ‘Trust, Economic Rationality, and the Corporate Fiduciary Obligation’, (1993) 43 *Univ Toronto LJ* 547, 550

⁹⁶ H. Collins, *The Law of Contract* (3rd ed, 1997) p 78.

⁹⁷ In particular Companies Act 1985, section 459.

⁹⁸ Evidence that shareholders in companies are insuring themselves against strategic behaviour are anecdotal but such a course has received judicial imprimatur and its appropriateness acknowledged by the Law Commission. See for instance *Cumbrian Newspapers Group Ltd v Cumberland & Westmorland Herald Newspaper & Printing Co* [1986] 3 WLR 26; G. Stedman and J. Jones, *Shareholder Agreements*, (2nd ed, 1990); *Russell v Northern Bank Development Bank Corp* [1992] 1 WLR 588.

⁹⁹ O-Kahn Freund “Some Reflections on Company Law Reform”, (1944) *MLR* 54,60; F. Fukuyama, *Trust: The Social Virtues and the Creation of Prosperity* (London, Hamish Hamilton, 1995) pp 3-4; K.A. Chrystal and R.G. Lipsey, *Economics for Business and Management* (Oxford, 1996) pp 123; J.H. Farrar & B. Hannigan, *Farrar’s Company Law* (4th ed, 1998) Ch 14-17.

which shareholders put up risk capital that is used to build the factory. . . and the job of management is to maximize profits for shareholders.”¹⁰⁰

It is this disembodied approach to wealth creation (for example, viewing capital as a *thing*) which is perpetuated by the uncompromising characteristics of majority rule, a bias towards the legal fiction of the company and the tacit acceptance of an adversarial approach to wealth creation. The evolution of company law and the attitude of its institutions towards managing severe divergences should not be restricted to its symbolic value. These developments can be seen as a deliberate move towards a more consensus oriented approach to wealth creation.¹⁰¹ In doing so, the spotlight is trained firmly on the *value of social relations* in the wealth creation process. Since economic relations between shareholders in closely-held firms are entrenched in relationships, it may seem foolish not to integrate this into the wealth creating process. It should be said that the idea of a set of social relations as generating value or ‘capital’ is not particularly controversial.¹⁰² There is less agreement on whether and to what extent legal instruments can be utilised to this end.¹⁰³ We are concerned here with values and norms like ‘trust’, ‘duty’ and ‘loyalty’ which enable parties to maximise the potential of their existing resources.¹⁰⁴ Unlike physical and human capital which are more tangible than the notion of social capital, this is not a basis for precluding the latter's capabilities in facilitating productive relationships, strengthening commitments in times of conflict and providing the impetus to attain particular ends or goals.¹⁰⁵ It is usual when discussing the value of ‘trust’ in the commercial environment to refer to a book written by Francis Fukuyama – *Trust: The Social Virtues and the Creation of Prosperity*.¹⁰⁶ In his comparison of the varied performances between economies, he observes the correspondence between high economic attainment and high levels of trust.¹⁰⁷ Economies in societies, he argues, where there is a trust deficit tend not to perform well.¹⁰⁸ One may perhaps not dissent from the underlying ideas but it does however require some obvious qualifications. For example, equating economic prosperity with a country's social capital is to take an over simplistic view of economic success or

¹⁰⁰ M. Blair, *Ownership and Control* (Washington, Brookings Institution) p 324.

¹⁰¹ R. Hardin, “Trust” in *The New Palgrave Dictionary of Economics and the Law*, (ed P Newman) (1998) p 623.

¹⁰² J. Charkham, *Keeping Good Company: A Study of Corporate Governance in Five Countries* (Oxford: Clarendon Press) Chapters 1-3; T. Sheridan and N. Kendall, *Corporate Governance: An Action Plan for Profitability and Business Success* (London: Pitman, 1992) Chapters 4-6 and 8.

¹⁰³ B. Cheffin, ‘Trust, Loyalty and Cooperation in the Business Community: Is Regulation Required?’ in B. Rider (ed), *et al The Realm of Company Law* (Oxford: Clarendon Press,) p 53.

¹⁰⁴ *Ibid* pp 53, 65-66, 70-78.

¹⁰⁵ J. Coleman, *Foundations of Social Theory*, (Cambridge, Harvard UP, 1990) pp 300-321.

¹⁰⁶ (London: Hamish Hamilton, 1995).

¹⁰⁷ *Trust: The Social Virtues and the Creation of Prosperity* (London: Hamish Hamilton, 1995)

¹⁰⁸ *Ibid* p 4; R. Hardin, “Trust” in *The New Palgrave Dictionary of Economics and the Law*, (ed P Newman) (1998) p 623.

underperformance. Comparisons between high-trust societies like Japan and one which has since ceased to possess such reserves like America, illustrate the dangers of adopting the thesis uncritically.¹⁰⁹ Another example can be seen in the world-wide financial crisis of 1997. This serves as a reminder of the chaos that can be caused by the absence of real constraints on the mobility of capital in the modern global market.¹¹⁰ The problems which originated in Thailand in July of that year rapidly spread across the globe dampening the economic prospects of the ‘Tiger Economies’, causing meltdown in Russia and the depreciation of the currencies in Latin America. Finally, economic prosperity in the age of technology depends on the ability of a business to gain a competitive advantage in the marketplace. It is generally recognised that success or failure of an enterprise cannot be confined to using social capital as a yardstick; much will depend on the entrepreneur’s ability to equip the business with organisational structures and processes which are both innovative as well as responsive to the pace and complexity of this commercial environment. Failing or underperforming economies / businesses in a number of instances have too often exhibited an ignorance of the evolutionary nature of capitalism and are unable to maximise the economic potential of the ‘technological revolution’.¹¹¹ Despite these reservations the thesis does contain an important message – it emphasises the *value* of social capital. James Coleman argues that the various elements which broadly constitute social capital – trust, duty and loyalty – enable it to perform an important function.¹¹² Where such elements are deemed to exist in abundance, individuals will be able to utilise these elements as resources to maximise their interests. It has already been shown in the discussion employing game strategy, how the absence of norms and values like ‘trust’ and ‘loyalty’ can undermine the efficacy of relations and the wealth creating potential of the commercial venture.¹¹³ As Fukuyama observes, a model which integrates social capital within the economic framework of the business enterprise will be more likely to facilitate productive relationships which have a “large and measurable economic value.”¹¹⁴

It is in this respect that the statutory oppression remedies, the introduction of rebuttable presumptions in specified situations and the availability of a statutory derivative action can be seen as an attempt to combine statutory instruments with the model of social relations to enable the realisation of the inherent potential of social resources. The legal instruments and proposed reforms attempt to minimise the deficiencies embedded in the contractual

¹⁰⁹ F. Fukuyama, *Trust: The Social Virtues and the Creation of Prosperity* (London, Hamish Hamilton, 1995) pp 29-30, 277-281, 310-311.

¹¹⁰ G. Soros, *The Crisis of Global Capitalism* (London, Little, Brown and Co 1998).

¹¹¹ J. Gray, *False Dawn* (London: Granta Publications, 1998); W Hutton, *The State We’re In* (London: Vintage 1996); G. Soros, *The Crisis of Global Capitalism* (London: Little, Brown and Co 1998)

¹¹² J. Coleman, *Foundations of Social Theory*, (Cambridge, Harvard UP, 1990) pp 301-306.

¹¹³ B. Cheffin, ‘Trust, Loyalty and Cooperation in the Business Community: Is Regulation Required?’ in B. Rider (ed), *et al, The Realm of Company Law* (Oxford: Clarendon Press) pp 53, 65-66, 70-78

¹¹⁴ F. Fukuyama, *Trust: The Social Virtues and the Creation of Prosperity* (London: Hamish Hamilton, 1995) p 10.

model of social relations by increasing the likelihood of convergences in interests, as well creating the conditions which reduce occasions for engaging in strategic or opportunistic behaviour.¹¹⁵

VII. CONCLUSION

Economic activity is founded on the notion that rational individuals enter into relations with a view to gaining from the co-operative venture.¹¹⁶ By and large relationships within companies are undertaken without recourse to the law.¹¹⁷ In view of this, one may be tempted to suggest that:

“ . . . business functions best when those engaging in commercial activity share a co-operative ethos and have a strong sense of trust and loyalty. . . various disadvantages of regulation indicate that it is undesirable to have obligations to act in a trustworthy, loyal and co-operative manner imposed by legislators or judges. Instead, to the extent that people rely on trust and related concepts when they carry on commercial activity, this should occur because that is how they believe they should behave, not because someone has told them to do so.”¹¹⁸

Whilst there is some merit in this view, it should be remembered that until the intervention of statute, company law was preoccupied with the market conception of the ‘company’; the deference to contractual ideology and the assumption of rational behaviour overlooked a fundamental tenet of the human condition, namely that:

“human agents who populate the economic institutions of capitalism are lacking in compassion.”¹¹⁹

When constructing the model of social relations, the legislators and judges seem to have either underestimated individuals’ incentives to act responsibly or the market’s capacity to discharge the function of distributive justice.¹²⁰ The irony, as the law continues to evolve in this regard, is that:

“ . . . the private and highly localized virtues of loyalty and trust within an institution like the corporation, at least as much as the unconstrained pursuit of self-interest through contracts, can

¹¹⁵ Bruce L Hay and Kathryn E Spier, ‘Settlement of Litigation’ in P. Newman (ed) *The New Palgrave Dictionary of Economics and the Law*, (1998) 442, 447.

¹¹⁶ M. Olson, *The Logic of Collective Action* (Cambridge: Harvard Univ Press, 1965) pp 1-3

¹¹⁷ *The Law Society, Memorandum by the Society’s Standing Committee on Company Law: A New Form of Incorporation for Small Firms* (The Law Society, No 79, 1981); *Forum of Private Business, A Report into Business Legal Structures* (London: 1991), *Company Law Review: The Law Applicable to Private Companies* (London: DTI, URN 94/529, 1994).

¹¹⁸ B. Cheffin, ‘Trust, Loyalty and Cooperation in the Business Community: Is Regulation Required?’ in B. Rider (ed) *et al, The Realm of Company Law* (Oxford: Clarendon Press) p 53, 79.

¹¹⁹ O.E. Williamson, *The Economic Institutions of Capitalism* (New York, 1985) p 391.

¹²⁰ M.J. Whincop, ‘Overcoming Corporate Law: Instrumentalism, Pragmatism and the Separate Legal Entity Concept’, (1997) 15 *Comp & Sec LJ* 411, 412.

also add up to the unconscious attainment of a greater good for all.¹²¹

¹²¹ B. Chapman, 'Trust, Economic Rationality, and the Corporate Fiduciary Obligation', (1993) 43*Univ Toronto LJ* 547, 588.

ANONYMOUS WITNESSES*

*Ruth Costigan, Lecturer in Law, University of Wales Swansea and
Professor Philip A Thomas, Cardiff Law School*

Seasoned observers of Northern Ireland's criminal justice system are familiar with a pattern of restrictive laws migrating to the legal system of England and Wales after a process of "normalisation"¹ initiated in Northern Ireland. So it is with the granting of anonymity to witnesses in criminal trials. The second inquiry into Bloody Sunday has brought the reality of witness anonymity to public notice.² Yet what has remained hidden is the increasing use of anonymous witnesses in criminal trials, a practice which has its roots in an exceptional trial in Northern Ireland but which has progressed swiftly through the process of normalisation so that it now affects routine prosecutions throughout the United Kingdom. This article traces the development and the expanding usage of anonymous witnesses and questions both the validity and appropriateness of the increasing acceptance of witness anonymity within the criminal justice system.

Witness intimidation is a major issue for agencies concerned with law enforcement and criminal justice.³ The public interest in the prosecution of offences and the interests of those affected by intimidation require a response to actual and feared⁴ retribution. Indeed, the Human Rights Act 1998 will enable vulnerable witnesses to raise Convention rights, notably those

* Grateful thanks to Helen Power (University of Glamorgan Law School), Penny Smith (Cardiff Law School) and David Campbell (Cardiff Law School) for their valuable comments on the draft.

¹ For discussion of the process of normalising the abnormal, see J Sim and P A Thomas, "The Prevention of Terrorism Act: Normalising the Politics of Repression" (1983) *JLS* 71.

² See B Hadfield, "*R v Lord Saville of Newdigate, ex p anonymous soldiers: What is the Purpose of a Tribunal of Inquiry?*" [1999] *PL* 663.

³ See for example, *Speaking up for Justice: Report of the Interdepartmental Working Group on the Treatment of Vulnerable or Intimidated Witnesses in the Criminal Justice System* (1998, Home Office); *The Statement of National Standards of Witness Care in the Criminal Justice System* (1996, Home Office); Law Commission Report No. 245, *Hearsay and Related Topics* (1997); *The Stephen Lawrence Inquiry. Report of an Inquiry by Sir William Macpherson*, Cm 4262 – I, (1999); *Masefield's Progress: How Heavy is the Administrative Burden on the Police Now?* HM Inspectorate of Constabulary Report (1997). A specific offence of witness intimidation exists in s 51, Criminal Justice and Public Order Act 1994 (as amended by Youth Justice and Criminal Evidence Act 1999, sch 4). The Youth Justice and Criminal Evidence Act 1999 contains a wide range of witness protection measures at the trial court's disposal. In 1996 Irene Adams' unsuccessful Witness Protection Bill proposed a statutory power to grant long-term anonymity to witnesses of organised violence or drug-related offences.

⁴ Unsurprisingly, witnesses' fear is disproportionate to the incidence of reprisals: W Maynard, *Witness Intimidation: Strategies for Prevention*, Police Research Group, Crime Detection and Prevention Series Paper No. 55 (1994; London: Home Office).

provided in Articles 2 (the right to protection of life by law) and 8 (the right to respect for private and family life), in support of a claim for protection.⁵ It is vital, nonetheless, to remain vigilant of the accused's right to a fair trial, which we contend is threatened by the current law and practice relating to witness anonymity.⁶ The first section of this article introduces the consequences of anonymity for fundamental principles of criminal justice. Section two considers the enhanced status conferred on the witness by anonymity and draws on popular culture for insights into this process. We then examine in detail, in sections three and four, the evolution of the anonymous witness, from judicial hearings involving agents of the state to commonplace trials with unidentified civilian witnesses. Finally, we assess the compatibility of domestic law with Article 6 of the European Convention for the Protection of Human Rights and Fundamental Freedoms (hereafter "ECHR"), which provides a number of rights to ensure a fair trial.

1. JUSTICE BLINDFOLDED

A fundamental feature of British justice is that its courts, procedures and participants are open to public scrutiny. Practitioners even in the most challenging of cases remain committed to it.⁷ This tradition has been described as Britain's "most enduring contribution to the law of other nations."⁸ It is not only justice itself which is served by openness: the maintenance of public confidence is dependent upon the process being subject to scrutiny by the individual and the media. Secret trials are anti-democratic and abusive. Only in exceptional circumstances and if the overall interests of justice demand it may this principle be set aside:

"[T]he broad principle is that the courts of this country must . . . administer justice in public. [But] as the paramount object must always be to do justice, the general rule as to publicity. . . must accordingly yield. But the burden lies on those seeking to displace its application in the particular case to make out that the ordinary rule must as of necessity be superseded by this paramount consideration. . . The. . . [judge] must treat [the question] as one of principle, and as turning, not on convenience, but on necessity. . . He who maintains that by no other means than by such a hearing can justice be done may apply for an unusual procedure. But he must make out his case strictly and bring it up to the standard which the underlying principle requires."⁹

⁵ See, for example, *Ludi v Switzerland* (1993) 15 EHRR 173; *Doorson v Netherlands* (1996) 22 EHRR 330.

⁶ See also, G Marcus, "Secret Witnesses" [1990] *PL* 207.

⁷ See for example, J Temkin, "Prosecuting and Defending Rape: Perspectives from Inside and Outside the Bar" (2000) *JLS* 219. Professor Temkin interviewed thirteen barristers with extensive experience of rape trials and found that they viewed the use of screens as contrary to open justice.

⁸ G Robertson, *Freedom, the Individual and the Law* (7th ed, 1993), p 345.

⁹ *Scott v Scott* [1913] AC 417, Viscount Haldane LC at 437–38.

In *Attorney General v Leveller Magazine*,¹⁰ Lord Diplock affirmed that, “The English system of administering justice does require that . . . in criminal cases. . . all evidence communicated to the court is communicated publicly.”

¹¹ This particular aspect of open justice is vital, for as Wigmore observed, public testimony produces:

“In the witness’ mind a disinclination to falsify; first, by stimulating the instinctive responsibility to public opinion, symbolised in the audience, and ready to scorn a demonstrated liar; and next, by inducing the fear of exposure of subsequent falsities through disclosure by informed persons who may chance to be present or to hear of the testimony from others present.”¹²

The use of anonymous witnesses in criminal trials endangers the fair administration of justice: at a general level, in the challenge to open justice, and, more specifically, in relation to the legal rights of the accused. Of prime importance is the defence’s Hohfeldian¹³ claim-right to examine the credibility and reliability of prosecution witnesses, termed the “right of confrontation”. This right is enshrined in United Kingdom jurisprudence,¹⁴ Article 6 of the ECHR,¹⁵ Article 14 of the International Covenant on Civil and Political Rights¹⁶ and in the Sixth Amendment to the United States Constitution.¹⁷ The right of confrontation is based on the need to “constitutionalise a barrier against flagrant abuses, trials by anonymous accusers, and absentee witnesses.”¹⁸ Three main aims are served by the right: the facilitation of cross-examination (the “greatest legal engine ever invented for the discovery of truth”);¹⁹ the presence in the courtroom of witnesses, enabling the tribunal of fact to assess credibility through observation of demeanour; and the provision of testimony on oath, guarding against falsehood. Where the right of confrontation is undermined, the presumption

¹⁰ [1979] AC 440.

¹¹ *Ibid*, at 449-50.

¹² J H Wigmore, *Evidence* (Chadbourn revision, 1976) vol 6, 435-36.

¹³ W N Hohfeld, “Some Fundamental Concepts as Applied in Judicial Reasoning” (1913) 23 *Yale LJ* 16; *Fundamental Legal Conceptions* (Yale University Press, 1919). In Hohfeld’s analysis, a claim-right is legally enforceable and imposes a correlative duty.

¹⁴ See, for example, *R v Radak and others* (1998) *The Times*, 7 October (CA).

¹⁵ Article 6(1): “In the determination of his civil rights and obligations or of any criminal charge against him, everyone is entitled to a fair and public hearing within a reasonable time by an independent and impartial tribunal established by law. . . (3) Everyone charged with a criminal offence has the following minimum rights. . . (d) to examine or have examined witnesses against him and to obtain the attendance and examination of witnesses on his behalf under the same conditions as witnesses against him. . .”.

¹⁶ The relevant provisions essentially reflect the language of Article 6 of the Convention, *supra*.

¹⁷ “In all criminal prosecutions, the accused shall enjoy the right to a speedy and public trial, by an impartial jury. . . and to be informed of the nature and cause of the accusation; [and] to be confronted with the witnesses against him. . .”.

¹⁸ *California v Green* 399 US 149 [1970] at 179.

¹⁹ *Ibid*, at 158, citing 5 Wigmore § 1367.

of innocence and the principle of full proof (reflected in the standard of proof beyond reasonable doubt) are also threatened.

In the South African case of *S v Leepile* (1985),²⁰ Ackermann J identified the principal restrictions on the defence where anonymous witnesses are employed:

“(a) No investigation could be conducted by the accused’s legal representatives into the witness’ background to ascertain whether he has a general reputation for untruthfulness, whether he has made previous inconsistent statements nor to investigate other matters which might be relevant to his credibility in general. (b) It would make it more difficult to make enquiries to establish that the witness was not at places on the occasions mentioned by him. (c) It would further heighten the witness’ sense of impregnability and increase the temptation to falsify or exaggerate.”²¹

This statement encapsulates the importance of the accused’s right to cross-examine prosecution witnesses and thereby challenge their credibility. The Supreme Court of the United States has also recognised the importance of the right to explore a witness’ background:

“When the credibility of a witness is in issue, the very starting point in ‘exposing falsehood and bringing out the truth’ through cross-examination must necessarily be to ask the witness who he is and where he lives. The witness’ name and address open countless avenues of in-court examination and out-of-court investigation. To forbid this most rudimentary inquiry at the threshold is effectively to emasculate the right of cross-examination itself. . . . Prejudice ensues from a denial of the opportunity to place the witness in his proper setting and put the weight of his testimony and his credibility to a test, without which the jury cannot fairly appraise them.”²²

Where anonymous witnesses are used, the defence is disadvantaged not only by severe limitations on strategy; it is also prejudiced by the symbolism of unidentified, physically screened witnesses. In addition to the elevation of the witness’ status (discussed below) there is the risk that the jury might infer that the defendant is a dangerous person, thus undermining the presumption of innocence. The use of screens is appropriate in some cases, particularly those involving charges of child abuse, but such practice causes particular difficulty where the protected witness is also anonymous. The leading case in England and Wales on the use of screens is *R v XYZ* (1990),²³ in which the Court of Appeal sanctioned the trial judge’s decision to protect child witnesses from the gaze of the defendants who were charged with sexually

²⁰ *S v Leepile and Others* (5) (1986) 4 SA 187.

²¹ *Ibid*, at 189.

²² *Smith v State of Illinois* 390 US 129 [1968], at 131–32, citing *Pointer v Texas* 380 US 404 [1965] and *Alford v United States* 282 US 687 [1931] respectively.

²³ (1990) 91 Cr App R 36.

abusing them.²⁴ In response to the claim that the screens risked prejudicing the jury, the appellate court commented: “We do not think, even without the warning which the learned judge did give to the jury, that any sensible jury could have been prejudiced against the defendant by the existence of this barrier between the witnesses and the dock.”²⁵ In *R v Schaub and Cooper* (1994),²⁶ the Court of Appeal considered the propriety of screening an adult witness (a rape victim) from the accused. In this situation, the Court was more cautious:

“There can be little doubt in our judgement that the use of screens is prejudicial to an accused person, even where the jury are properly warned not to make any assumptions adverse to the accused person because of the presence or use of screens. The very fact that they are being employed at all suggests to a jury that there is a need for the witness to be protected in some way from any contact, even if it is only visual, with the defendant. Accordingly, the defendant is to some extent at a disadvantage. In our judgement, it should only be in the most exceptional cases that apparatus of this kind should be used when an adult is giving evidence.”²⁷

Following the high-profile rape trials of Ralston Edwards and Milton Brown,²⁸ Parliament recognised the trauma experienced by adult witnesses in sexual abuse cases: the Youth Justice and Criminal Evidence Act 1999 provides for the screening of such witnesses and of other witnesses who might be afraid or distressed by the prospect of testifying.²⁹ However, the endowment of *anonymity* increases the prospect of prejudice aroused by the presence of screens, and as the appellate court has recognised, admonitions to the jury are of limited efficacy.

2. ANONYMOUS “SUPERWITNESSES”

The unidentified or screened witness is, and is perceived as, different from the named, visible witness. The special status enhances the deviation from normal practice in that signifiers upon which evidential evaluations are based are excluded. Michael Mansfield QC who has appeared as defence counsel in cases involving anonymous witnesses observed that, “The hidden witness does nothing to advance the cause of justice. Rather it encourages the development of value judgements about the witnesses’ credibility which are based on the granting of *anonymity* rather than on the evidence itself.”³⁰ This

²⁴ This was an exercise of the trial judge’s common law discretion in furtherance of his duty to ensure a fair trial, consequently no authority was needed to justify his decision. The use of screens in cases involving children has in many cases been obviated by the statutory power to receive a child’s evidence given remotely via a live video-link (Criminal Justice Act 1988, s 32 as amended by Criminal Justice Act 1991, s 55).

²⁵ (1990) 91 Cr App R 36, at 40.

²⁶ [1994] Crim LR 531.

²⁷ *R v Schaub and Cooper*, CA, 29 November 1993; transcript.

²⁸ (1998) *The Times*, 7 May.

²⁹ Ss 17, 23; the Act also prohibits cross-examination in person by those charged with sexual offences (s 34).

³⁰ Interview with Philip Thomas, 1998.

observation is supported by Jane Winter, director of British Irish Rights Watch: “In the course of sending independent observers to trials and inquests in Northern Ireland over the past seven years it has been our experience that screening and anonymity, usually linked to public interest immunity certificates, have been increasingly used to hide the identity of key police and army witnesses. In our view, such devices do nothing to assist in establishing the facts of the case or delivering a fair trial.”³¹

We constantly make judgements, silent or otherwise, about people. Police officers, for instance, have to make instantaneous decisions concerning appropriate responses in fast changing public situations. Skolnick observed that officers develop profiles of “symbolic assailants” who might, for example, wear leather jackets, motor bike boots and “strut” around. Having identified the potential aggressor the interactive behaviour is modified to respond accordingly.³² Similarly, MacInnes argues that police officers have a profound dislike of “people loitering in streets, dressing extravagantly, speaking with exotic accents, being strange, weak, eccentric, or simply [appearing to belong to] any rare minority – of their doing, in fact, anything that cannot be safely predicted.”³³ The power of the stereotype to influence judgements about criminality was notoriously demonstrated in the work of Cesare Lombroso (1836-1909), who converted stereotypes into “scientific facts.” His standard criminal appearance reflected prominent rodent-like incisors, a receding or flat chin with little or no beard, bushy eyebrows which either joined or had a devilish curl at the outer ends. Murderers often had dark rather than blond hair and fraudsters tended towards curly hair. Whilst Lombroso’s “science” has been discredited, the power of the stereotype continues:

“How can you possibly say that a person looks like a criminal? Several policemen talk about the ability to ‘feel’ or ‘smell’ a criminal, or to have a ‘sixth sense’ about a person. What they are really talking about is the ability to see a criminal when they come across one. It is basically a question of being able to categorise or stereotype a person . . . the most skilful policeman will therefore not only be able to recognise a criminal when he sees one but will often be able to state what type of previous convictions the particular criminal has.”³⁴

³¹ Private correspondence with the authors, 1998.

³² J H Skolnick, *Justice Without Trial: Law Enforcement in Democratic Society* (1966).

³³ *Mr Love and Justice* (1962).

³⁴ J Robinson, *Catching Criminals: Some Basic Skills* (1978). Several studies suggest that facial stereotypes exist for various types of criminals. See for example, G R Thornton, “The Ability to Judge Crimes from Photographs of Criminals” (1939) *J of Abnormal and Social Psychology* 378-83; R H C Bull and J Green, “The Relationship between Physical Appearance and Criminality” (1980) *Medical Science Law* 79-83; A G Goldstein, J E Chance & B Gilbert, “Facial Stereotypes of Good Guys and Bad Guys” 22 *Bulletin of Psychonomic Society*, 549-52. For a general account of crime and facial appearance see R Bull and N Rumsey, *The Social Psychology of Facial Appearance*, 1988, chapter 4: “Facial appearance and the criminal justice system” 81-101.

Judging the individual is an action members of the jury have undertaken on countless social occasions. But doing so in court is a new experience, subject to different rules and a restricted set of indicators. The formal and social rules that operate within the court privilege the regular actors: the professionals, the court staff and the police. An arcane linguistic code is employed to control the involvement of the witness and the defendant.³⁵ Features such as dialect,³⁶ appearance,³⁷ ethnicity,³⁸ gender, status and demeanour affect witness credibility. The United States Supreme Court has explained that jurors must be able to see the witness, so that they may “judge his demeanour upon the stand and the manner in which he gives his testimony whether he is worthy of belief.”³⁹ In *Coy v Iowa* (1988),⁴⁰ the Court observed that the right of confrontation is guaranteed the defendant “because it is always more difficult to tell a lie about a person to his face than behind his back.”⁴¹ As Jerome Frank commented,

“All of us know that, in everyday life, the way a man behaves when he tells a story – his intonation, his fidgeting or composure, his yawns, the use of his eyes, his air of candour or of evasiveness – may furnish valuable clues to his reliability. Such clues are by no means impeccable guides, but are often immeasurably helpful. So the courts have concluded.”⁴²

Shepherd, a forensic psychologist and psychotherapist, has noted that, “judgements of initial credibility arise from the manner in which the oath is taken, comportment and the representation of reality (material facts) and non-verbal cues exhibited when giving this through the process of examination-in-chief.” When evaluating witnesses, jurors use “four key elements: language, intonation, sounds outside language and, finally, body language such as posture, facial expression, eye contact and gaze.”⁴³

The face is widely recognised as the most important area of our bodies in influencing and regulating personal interactions;⁴⁴ it is the principal area on which the viewer concentrates for non-verbal communication. For example, if actors, poker players and politicians were unable to control or manipulate their faces their status, income and credibility would be severely affected. Identifying the lie is difficult but it is a popular belief that the face is a mirror, albeit sometimes cloudy, which reflects the honesty of the individual. However, research suggests that exposure to liars does not produce greater

³⁵ See, for example, P Carlen, “Remedial Routines for the Maintenance of Control in Magistrates’ Courts” (1974) *BJLS* 101.

³⁶ See W M O’Barr, *Linguistic Evidence: Language, Power and Strategy in the Courtroom* (1982); Hans and Vidmar, *Judging the Jury* (1986).

³⁷ See, for instance, J Temkin, *op cit* n 7.

³⁸ See Hans and Vidmar, *Judging the Jury* (1986).

³⁹ *Mattox v US* 156 US 237, 242-43 (1895).

⁴⁰ *Coy v Iowa* 487 US 1012 (1988).

⁴¹ *Ibid*, at 1016, 1019.

⁴² J Frank, *Courts On Trial* (1950), p 21.

⁴³ Taken from an opinion from Dr Shepherd, generously provided by Bindman and Partners, London (1999).

⁴⁴ See generally, T R Alley (ed), *Social and Applied Aspects of Perceiving Faces* (1988); J Liggett, *The Human Face* (1974); F C McGregor, *Transformation and Identity* (1974).

detection acuity, even amongst criminal investigators, polygraphers, judges and psychiatrists.⁴⁵ In view of this and of the dangerously seductive power of the stereotype, it might be thought that the witness' visible presence in court could actually be detrimental to the discovery of the truth. However, we contend that as the whole process of trial is a heavily constructed piece of theatre, to draw a curtain over some of the characters and props paradoxically throws the spotlight on them, distorting their significance and affecting the plot. The combination of the story, the storytellers and the audience produces a complicated pattern within which credibility is established. Words themselves are only part of this process of defining accurate and credible testimony. There are so few Pinocchios that the court is obliged to read the face, observe the demeanour and hear the words of the witness.

The absence of name, occupation, address and, if screened, the person, concentrates the jury's attention on the voice, which might be subject to electronic distortion. The elements of mystery and importance are introduced and endorsed by the state. It is the state, through the prosecution, which claims that this person is so important or vulnerable that only limited access can be allowed to the jury, the defendant, counsel and the public. Such endorsed witnesses become "super-witnesses" whose credibility should neither be tested nor doubted. Who is this person, being so important, that he cannot be identified to the very court which is empowered to test the evidence? Disbelieving or doubting such a witness involves casting doubt on the state's evaluation procedures. For jurors this is a double burden to carry.

It is from literature and popular culture that we receive our directions as to what is good and bad, strong and weak, desirable and undesirable. Mass culture informs the jury as to witness credibility.⁴⁶ Insights into the status anonymity confers on a witness are to be gained from an examination of anonymity as a literary device. Absence of the main character in psychological thrillers is a common technique employed to heighten the character's menace, as exemplified by Patricia Cornwell's serial killer, Temple Gault, whose identity is only revealed to the reader at a late stage in the series of novels built around him. The identity of Alexander Dumas' "Man in the Iron Mask" became a public guessing game; it was simply too important and politically divisive to be revealed. "The Invisible Man" of H G Wells was fascinating because of his apparent lack of physical presence.⁴⁷ Superman, the Lone Ranger, Batman, Robin and Spiderman are, in their bizarre clothing and masked presence, infinitely more exciting than their known alter egos. Clark Kent was able to attract Lois Lane's sexual interest only in his Superman persona. In the 1940s Superman films, audiences were

⁴⁵ D McNeil, *The Face* (1998); P Ekman, *Telling Lies* (1986); O G Wellbourn, "Demeanour" (1991) 76 *Cornell LR* 1104; M Stone, "Instant Lie Detection? Demeanour and Credibility in Criminal Courts" [1991] *Crim LR* 821.

⁴⁶ See the influential work of H Innes, *The Bias of Communication* (1975) and M McLuhan, *Understanding Media, The Extensions of Man* (1964); *The Gutenberg Galaxy: The Making of Typographic Man* (1962).

⁴⁷ R F Fleissner, "H G Wells and Ralph Ellison: Need the effect of one invisible man on another be itself invisible?" (1992) 33 *Extrapolation* 346–50; A B Simpson, "The 'Tangible Antagonist': H G Wells and the Discourse of Otherness" (1990) 31 *Extrapolation* 134–47.

encouraged to believe that Superman was playing himself by the exclusion of Kirk Alyn's name from the credits.⁴⁸ Clayton Moore, who played the Lone Ranger, was contracted to remain publicly anonymous. Thus the role, not the actor, became the icon. Clint Eastwood's nameless, emotionless cowboy who featured in westerns such as "High Plains Drifter," offered no history and no future; he appeared, dispensed bloody justice and left. On stage, in "The Phantom of the Opera," Christine Daae seeks to snatch the mask from Erik's face; she cries that she "wants to know the face of the voice."

The sociologist, George Simmel, noted that every sense contributes to the messages we receive and thus to the judgements we make of others.⁴⁹ He believed that "the eye is destined for a completely unique sociological achievement: the connection and interaction of individuals. . . perhaps the most direct and purest interaction that exists."⁵⁰ Direct eye contact conveys considerable information, offering the observer a version of reality which might or might not corroborate the oral message. In the case of the anonymous witness, the defendant, jury, public and the media are deprived of this source of information. Furthermore, as Simmel observed, secrets adorn and enhance personality: the mask bestows instant mystique. So, paradoxically, the anonymous witness, the one about whom the jury knows the least, has the highest status.

3. ANONYMOUS STATE AGENTS

"The Tribunal has as its fundamental objective the finding of the truth about Bloody Sunday. It regards itself as under a duty to carry out its public investigative function in a way that demonstrates to all concerned that it is engaged in a thorough, open and complete search for the truth about Bloody Sunday."⁵¹ "[T]hese are the very soldiers whose conduct lies at the centre of this Inquiry. To allow this group to remain entirely anonymous would be a step that we would find difficult to reconcile with our public duty to determine what happened on Bloody Sunday."⁵²

As the above statement by the Saville Tribunal (appointed in 1998 to conduct a second inquiry into Bloody Sunday) indicates, the use of anonymous witnesses not only detracts from the appearance of openness but increases the prospect that justice will be obstructed. Despite these dangers, there has in recent years been a quantum leap in the granting of anonymity to state agents, be they police officers or members of the intelligence services, the SAS or the army (soldiers, for example, have routinely testified

⁴⁸ Alyn recalled, "Playing Superman ruined my acting career. I couldn't get another job in Hollywood." Obituary, *The Guardian*, 24 March 1999.

⁴⁹ See D Frisby and M Featherstone, *Simmel on Culture* (1997) specifically, "Sociology of the Senses" at p 109.

⁵⁰ *Ibid*, at 111.

⁵¹ Further Rulings by Lord Saville, Mr William Hoyt, Sir Edward Somers, 5 May 1999, para 11: <http://www.bloody-sunday-inquiry.org.uk>.

⁵² Further Rulings and Observations by Lord Saville, Mr William Hoyt, Sir Edward Somers, 17 December 1998, para 46: <http://www.bloody-sunday-inquiry.org.uk>.

anonymously at inquests in Northern Ireland).⁵³ The practice of granting anonymity to state agents in judicial hearings has led, through a process of normalisation, to an insidious increase in the use of anonymous civilian witnesses⁵⁴ in criminal trials. Before examining this development, it will be instructive to consider the cost to the fair administration of justice of endowing state agents with anonymity.

One of the most prominent illustrations of this form of secret justice was the inquest into the 1988 killing by British security forces of three unarmed members of the IRA in Gibraltar.⁵⁵ There was widespread concern over the deaths amid suggestions that Sean Savage, Mairead Farrell and Daniel McCann had been summarily executed.⁵⁶ The Home Secretary, Sir Geoffrey Howe attempted to prevent transmission of a television documentary, "Death on the Rock", which challenged the Government's account of the incident.⁵⁷ At the inquest in Gibraltar, eighteen witnesses from the SAS, M15, Special Branch and the local police testified anonymously and from behind a screen. Furthermore, counsel for the United Kingdom government, John Laws, invoked public interest immunity certificates to prevent questioning of witnesses which threatened "national security".

The coroner lacked the power to compel the SAS soldiers to testify and it was doubtful whether they would attend the inquest. Their testimony was eventually secured on assurances of anonymity and screening. The coroner, Felix Pizzarello, commented: "The reality seems to be that unless the witnesses are screened I may not have a meaningful inquest, and, of course, if they are screened it would be a flawed inquest in any case."⁵⁸ Journalists' accounts of the inquest reflected the procedural constraints:

"Soldier A was clearly working-class and from the South of England – perhaps London. This much could be deduced from his accent. He was the one who fired the first shot. . . Down the street were soldiers C and D. They were also working-class but from the north: soldier C was probably from Lancashire. . .

⁵³ For instance, *R v Newcastle upon Tyne Coroner, ex p. A* (1997, *The Times*, 19 December) involved a successful challenge to a coroner's refusal to allow a police officer to testify behind a screen at the inquest into the officer's fatal shooting of a burglar. The Divisional Court concerned itself solely with the issue of screens, neglecting to consider the consequences of the anonymity granted to officer "A". For illustrations of soldiers testifying anonymously, see *The Guardian*: 27 April 1992, 3 February 1993, 10 December 1993, 22 October 1997; *The Independent*, 7 October 1992.

⁵⁴ See S Enright, "The Anonymous Witness" (1996) *NLJ* 1032.

⁵⁵ The inquest jury returned a majority verdict of lawful killing but, in an unprecedented decision by the European Court of Human Rights, the UK was subsequently found in violation of Article 2 of the ECHR: *McCann and others v UK* (1995) 21 EHRR 97.

⁵⁶ Enoch Powell MP is reported as saying that the security forces might have committed "deliberate, cold-blooded, premeditated murder." *The Independent*, 1 April 1988.

⁵⁷ This programme and the surrounding events produced an inquiry which reported as *The Windlesham/Rampton Report on Death on the Rock* (1989).

⁵⁸ J Tweedie, *The Gibraltar Inquest Report* (1988) *Inquest* 5. See also, J Tweedie and T Ward "The Gibraltar Shootings and the Politics of Inquests" 16 (1989) *JLS* 464.

Officers E and F spoke as though they had attended public school. . . the final witnesses were policeman P, policeman Q and policeman R.”⁵⁹

Pizzarello referred to the importance of being able to assess the credibility of the witnesses by their physical demeanour, which in this hearing proved impossible. The soldiers claimed that the three deceased had made suspicious movements on being confronted, as if reaching for a remote detonator. These movements led directly to the shootings which were ostensibly to prevent the detonation of a suspected, but non-existent, car bomb. The screens prevented a comparison of the soldiers’ demonstrations with those of eyewitnesses. The soldiers who shot Savage claimed he was facing them and explained the bullet wounds in his back on the basis that he had “corkscrewed” to the ground. Because of the screens their demonstration of the fall was not visible to the public and the press. Whilst these features impacted on the watching world rather than on the jury, there were other illustrations of anonymity which affected the jurors. Photographs of people around the bodies, taken after the shooting, had been doctored; faces were obscured which resulted in civilian witnesses being unable to verify their own observations.⁶⁰

The authors interviewed a former SAS soldier with extensive experience in Northern Ireland.⁶¹ Having served with soldiers involved in the Gibraltar killings, he observed of that incident: “The soldiers felt pilloried after conducting an excellent operation. They were not happy about giving evidence even after an M15 briefing because of the controversy. They were concerned about being charged with murder. The anonymity was claimed for them although I suspect they would have claimed it anyway.” He had direct experience of this when he and four other soldiers, all of whom were involved in a surveillance operation in Northern Ireland, testified anonymously from behind a screen at a criminal trial. The decision to seek anonymity was taken by the army *without consultation* with the men: “It wouldn’t have worried me to give evidence in open court. I am big enough and ugly enough to look after myself. I wasn’t afraid. We all felt stupid standing behind a screen. We were professional, elite soldiers and we don’t scare easily.”

Some of the deficiencies of the “Gibraltar 3” inquest will be apparent in the second inquiry into Bloody Sunday. The Saville Tribunal twice decided⁶² to refuse anonymity to all military witnesses, including those soldiers who had admitted at the Widgery Inquiry⁶³ to firing live rounds during the civil rights march of 30 January 1972. Each of these decisions was successfully

⁵⁹ I Jack “Gibraltar” 25 *Granta* 1988, 43/44.

⁶⁰ *Ibid.*

⁶¹ The ex-soldier who had served as a regular soldier for fifteen years insisted on anonymity before undertaking the interview in 1998.

⁶² See <http://www.bloody-sunday-inquiry.org.uk>, rulings of 17 December 1998 and 5 May 1999.

⁶³ *Report of the Tribunal appointed to Enquire into the Events of Sunday 30th January 1972, which led to the Loss of Life in Connection with the Procession in Londonderry on that Day* HC 220, 1971-72, HMSO, London, 1972.

challenged by way of judicial review.⁶⁴ As a result, the Tribunal announced in July 1999 that it would not appeal further against the orders of certiorari. Accordingly, *all* soldiers present at Bloody Sunday will give evidence anonymously at the full public hearing which commenced in March 2000.⁶⁵ They will have the further protection of an assurance by the Attorney General that their evidence will not be used against them in any subsequent criminal proceedings.⁶⁶ There is much that could be said about the courts' rejection of the Tribunal's reasons for denying the soldiers anonymity, but of particular relevance here is the courts' approach to the issue of open justice.

The main aim of the second inquiry into Bloody Sunday is, in the Prime Minister's words, that "the truth be established and told."⁶⁷ The Widgery Inquiry was discredited, not only because it arguably failed to find the truth but also because its proceedings lacked transparency. As the lawyers for the family of James Wray, who was killed on Bloody Sunday, observed:

"One of the most objectionable features of Widgery was that soldiers gave their evidence under the cloak of anonymity, told lies and were never prosecuted or called to account. If . . . [the Saville] Inquiry adopts the same practice on anonymity, it will attract the same cynicism and disrespect. A process that starts by covering up the names of key witnesses will be seen as yet again covering up the truth. By contrast, an Inquiry which starts by making it clear that it proposes to conceal nothing of relevance and that it expects witnesses to come forward, identify themselves and be subjected to fair and public scrutiny, will deserve and will receive the confidence and participation of the public."⁶⁸

The Saville Tribunal initially ruled that the soldiers seeking anonymity bore the obligation of justifying their application,⁶⁹ noting that, "it is not open justice that needs to be justified, but rather any departure from open

⁶⁴ *R v Lord Saville of Newdigate and others, ex p B, O, U & V*, unreported, 16 March 1999 (DC) and (1999) *The Times*, 15 April (CA); *R v Lord Saville of Newdigate and others, ex p. A and others*, (1999) *The Times*, 22 June (DC) and 29 July (CA).

⁶⁵ In relation to the soldiers not accused of firing live rounds, the Court of Appeal acknowledged in its judgment of 28 July 1999 that it was for the Tribunal to decide whether they should be accorded anonymity, commenting only that on the material currently before it the Court could not state that a decision to refuse anonymity would be unlawful. In view of the criticism it had received in the Court of Appeal, the Tribunal felt that it had no choice but to extend anonymity to all soldiers, which it did in its ruling of 12 October 1999: <http://www.bloody-sunday-inquiry.org.uk>.

⁶⁶ This assurance does not, however, amount to an immunity against prosecution: Attorney General Press Notice, 25 February 1999: <http://www.bloody-sunday-inquiry.org.uk>.

⁶⁷ HC Hansard, 29 January 1998.

⁶⁸ Lord Gifford QC and Barry Macdonald, submission to the Bloody Sunday Tribunal, 12 November 1998: <http://www.bloody-sunday-inquiry.org.uk>.

⁶⁹ Rulings and Observations of the Tribunal on the Matters Raised at the Preliminary Hearing on 20th and 21st July 1998, published 24 July 1998: <http://www.bloody-sunday-inquiry.org.uk>.

justice.”⁷⁰ The Court of Appeal, however, expressly disapproved of this approach when reviewing the Tribunal’s subsequent decisions on anonymity. Lord Justice Otton, in the appellate hearing of March 1999, commented that the Tribunal members “might wish to reconsider the fairness of imposing the obligation ‘on those who seek anonymity of any kind to justify their claim’ . . . Similarly, they may wish to revisit their requirement that there must be ‘concrete evidence of specific threat’.”⁷¹ This view was endorsed by Lord Woolf MR in the final judicial review hearing:

“The Tribunal then state. . . that in their judgement ‘it is not open justice that needs to be justified but rather any departure from open justice.’ Again, this paragraph does appear to play down the significance which should be attached to the risk to the soldiers. Surely it could be said equally that the need for increasing the risk to the soldiers has to be justified.”⁷²

Indeed, the appellate court concluded that the Tribunal, having accepted that the soldiers had reasonable grounds to fear for their safety if identified, bore the obligation of finding a compelling justification for refusing anonymity.⁷³ This inversion of fundamental principle had previously been manifested by the Divisional Court: “by requiring the applicants to justify a departure from public and open justice. . . [the Tribunal] wrongly gave precedence to the consideration of the carrying out of a public investigation.”⁷⁴

In addition to departing from the traditional approach to open justice, the courts naively posited that “the open search for the truth would only be restricted in a marginal way”⁷⁵ by extending anonymity to the soldiers. This gives insufficient weight to the Tribunal’s observation that,

“the conduct of these soldiers lies at the very heart of this Inquiry. It is the firing on the streets that was the immediate cause of loss of life. It is that loss of life that we are publicly investigating. To conceal the identity of those soldiers would . . . make particularly significant inroads on the public nature of the Inquiry.”⁷⁶

Given the obvious importance of an open investigation, it is surprising that the Court of Appeal felt able to conclude so confidently that, “we do not

⁷⁰ Further Rulings by Lord Saville, Mr William Hoyt, Sir Edward Somers, 5 May 1999: <http://www.bloody-sunday-inquiry.org.uk>.

⁷¹ *R v Lord Saville of Newdigate and others, ex p B, O, U & V*, (1999) *The Times*, 15 April; transcript.

⁷² *R v Lord Saville of Newdigate and others, ex p A and others*, (1999) *The Times*, 29 July; transcript.

⁷³ *Ibid*, para 68 of transcript.

⁷⁴ *R v Lord Saville of Newdigate and others, ex p A and others* (1999) *The Times*, 22 June.

⁷⁵ *R v Lord Saville of Newdigate and others, ex p A and others*, (1999) *The Times*, 29 July; para 56 of transcript, Lord Woolf MR. See also the judgment of Roch LJ in *R v Lord Saville of Newdigate and others, ex p A and others* (1999) *The Times*, 22 June (DC).

⁷⁶ *Op cit* n 70, para 28.

consider that any decision was possible other than to grant the anonymity to the soldiers.”⁷⁷

When state agents testify anonymously in criminal trials, it is not only open justice which is at stake: so too are the accused’s legal rights. The courts have justified this practice on the basis of protecting the witnesses from physical harm and of preserving their future operational usefulness. Yet the necessity of withholding these witnesses’ identities often remains unestablished. Perhaps the clearest illustration of this is *R v Roberts, Davies and Williams* (1993),⁷⁸ the trial of Welsh nationalists for the possession and use of explosives and conspiracy to cause explosions. MI5 had employed 38 officers during covert surveillance of political rallies and of a residence in Anglesey. At trial, Pill J rejected the prosecution’s application for the proceedings to be held *in camera*, but did grant anonymity to security service witnesses. The officers testified as Mr A, B, C, D, E and F from behind a screen. However, outside the court these officers were easily identified as they stayed in one hotel, kept to themselves and spoke English in a predominantly Welsh speaking community.

The interests of “national security” have also been employed to justify anonymity. The Court of Appeal in *R v Jack* (1998),⁷⁹ an appeal against conviction and sentence for conspiracy to cause explosions, denied that national security was used as a trump card to legitimate the anonymity afforded fourteen Security Service witnesses: “The security services do not have any passport to the use of screens, or to any other form of anonymity. Their position has to be considered as individual witnesses on a case by case basis.” An analysis of this case indicates that courts might in future seek to avoid the traditional charge of submissiveness in the face of “national security” claims by justifying abnormal administration of criminal justice on the basis of intelligence agents’ fear of reprisals. This approach will, despite the Court of Appeal’s protestation in *Jack*, facilitate the granting of anonymity to undercover police and MI5 officers, an undesirable development particularly in view of the extension of MI5’s role into the investigation of “serious crime.”⁸⁰

At the trial of Jack and his co-defendant, Fryers, the crown sought anonymity for Security Service officers who had engaged in covert surveillance of the two men. The application was supported by a public interest immunity certificate which, the Court of Appeal noted, was “couched in very general terms.” The trial judge, after balancing the interest of open justice against that of national security (in accordance with the traditional approach),⁸¹

⁷⁷ *R v Lord Saville of Newdigate and others, ex p A and others*, (1999) *The Times*, 29 July; para 69 of transcript, Lord Woolf MR.

⁷⁸ Unreported. See P A Thomas “Secret Police on Trial” 98 (1993) *Planet: The Welsh Internationalist*, 3-8 for an account of the trial and the surrounding politics.

⁷⁹ *R v Jack (Hugh Thomas)*, CA, 7 April 1998, unreported; extracts reproduced from transcript.

⁸⁰ Security Services Act 1989, s 1(4), inserted by the Security Services Act 1996, s 1.

⁸¹ This approach is exemplified in the words of Hutton LCJ: “Where the claim of national security is raised in such a way that it may conflict with the principle of open justice, the court must balance the claims and decide whether there should be a restriction of the principle of open justice in order to meet the claims of national

granted the request and the witnesses, each identified by a letter, gave their testimony screened from the accused, the press and the public gallery. On appeal, counsel for Jack contested the use of the screens (though not the employment of letters to denote each witness), arguing that the trial judge had given too much weight to the fact that MI5 agents were involved and had paid insufficient regard to the defendant's interests and to the prejudice aroused by the nature of the charges (the prosecution alleged that Jack had acted for the IRA). Jack's counsel, Ben Emmerson, also cautioned the Court about the threat to normal trial procedures posed by the increasing claims to anonymity made by Security Service personnel.

Whilst upholding the trial judge's decision to grant anonymity, the Court of Appeal redefined the balancing exercise to be undertaken in national security cases. The Court, preferring not to pursue arguments based on principles of public interest immunity, indicated that a departure from the principle of open justice may be sanctioned only by reference to the administration of justice and not in furtherance of other public interests. The Court explained that,

“On that basis the balance to be struck is between two aspects of the administration of justice and it is not permissible to balance national security *per se* against the administration of justice. Statute apart, considerations of national security can justify a departure from the principle of open justice only so far as they have an effect upon the administration of justice itself, *e.g.* by deterring the crown from continuing a prosecution, or, say, by deterring an individual member of the Security Services from giving evidence out of fear for his own safety.”⁸²

Although superficially attractive, this decision discourages an in-depth judicial examination of what is required in the interests of national security; the focus on a witness' fear and on the crown's reluctance to call a witness denied anonymity will, it is contended, lead to an increase in the use of anonymous law enforcement witnesses. Further, the Court's ostensible insistence on undercover agents being treated as ordinary individuals when a claim for anonymity is made is at variance with recent ECHR jurisprudence. As the European Court of Human Rights observed in *Van Mechelen v The Netherlands*,⁸³

“The balancing of interests of the defence against arguments in favour of maintaining the anonymity of witnesses raises special problems if the witnesses in question are members of the police force of the State. Although their interests – and

security.” *R v Coroner for Greater Belfast*, Court of Appeal in Northern Ireland, 22 April 1993, unreported.

⁸² The Court of Appeal did not, however, preclude the possibility of anonymity being justified on the basis of public interest immunity. As the Court concluded that in *Jack* the considerations of national security and of the administration of justice “merged into one”, it left open the question whether a departure from open justice could be justified on the basis of national security issues which had no impact on the administration of justice.

⁸³ (1998) 25 EHRR 657.

indeed those of their families – also deserve protection under the Convention, it must be recognised that their position is to some extent different from that of a disinterested witness or victim. They owe a general duty of obedience to the State's executive authorities and usually have links with the prosecution; for these reasons alone their use as anonymous witnesses should be resorted to only in exceptional circumstances. In addition, it is in the nature of things that their duties. . . may involve giving evidence in open court.”⁸⁴

Furthermore, ECHR jurisprudence requires that where undercover agents are afforded anonymity as witnesses, the rights of the defence be respected.⁸⁵ In the *Jack* case, the prosecution evidence consisted mainly of surveillance reports, yet the defence was unable to investigate the credibility of the anonymous witnesses and was prevented from ascertaining the location of a surveillance camera and the capacity of tracking devices employed to monitor Jack. The appellate court rejected the submission that these circumstances denied counsel the opportunity to present an effective defence. The limitations imposed on the defence by the use of anonymous witnesses were not fully canvassed in the Court of Appeal, but in relation to the screening of witnesses the Court made the remarkable statement that,

“Where the charge under consideration is a terrorist offence the prejudice to the accused of allowing screens to be used is likely to be somewhat reduced, because the jury may have to have protection themselves, and will easily appreciate why, if the charge is well-founded, protection should be afforded to a witness.”

A fundamental objection to the courts' approach to cases involving state agents is that they do not entertain the possibility of the intelligence agencies, rather than the defendant, bearing the burden of any compromise between the prosecution of offences and operational efficacy. The judiciary could adopt the position that due process should not be sacrificed for the convenience of the intelligence services, which would have to accept that the price of a prosecution might be the (temporary) transfer or limitation of specific officers' activities. This might prove a burden to law enforcement agencies but it is arguably preferable to the denial of fairly administered criminal justice. In relation to the justification based on protection from physical harm, consideration should be given, where appropriate, to measures short of anonymity (such as press reporting restrictions and other forms of witness protection), which would be more compatible with the accused's right of confrontation. The granting of anonymity to state agents should be on the basis of necessity, rather than convenience, with the court's decision being made on the provision of evidence as to the level of risk to each individual seeking such protection.

⁸⁴ *Ibid*, para 56.

⁸⁵ *Van Mechelen op cit* n 83; *Ludi v Switzerland op cit* n 5; *Kostovski v The Netherlands* (1989) 12 EHRR 434; *Doorson v The Netherlands op cit* n 5.

4. THE NORMALISATION OF ANONYMITY

“The need to be vigilant arises from the natural tendency for the general principle to be eroded and for exceptions to grow by accretion. . . . This is the reason it is so important not to forget why proceedings are required to be subjected to the full glare of a public hearing. It is necessary because the public nature of proceedings deters inappropriate behaviour on the part of the court. It also maintains the public’s confidence in the administration of justice. It enables the public to know that justice is being administered impartially. It can result in evidence becoming available which would not normally become available if the proceedings were conducted behind closed doors or with one or more of the parties’ or witnesses’ identity concealed.”⁸⁶

The extraordinary is becoming the ordinary; the evolution of the anonymous witness can be traced from judicial inquiries into killings by state agents to committal proceedings for violent disorder. Yet the increasing use of anonymous witnesses is rooted in an inappropriate extension of the common law. Two of the cases employed by courts as authority for witness anonymity do not in fact bear that interpretation. *R v Socialist Worker, ex parte Attorney General* (1974)⁸⁷ was a contempt case arising from the publication of the names of two prosecution witnesses in a blackmail trial. The witnesses had testified in open court but their names were concealed. At the contempt hearing, it was argued on behalf of *Socialist Worker* that the trial judge lacked the power to conceal witnesses’ names (thus the publication could not be contemptuous). The Divisional Court, however, referred to the long-established practice of allowing blackmail witnesses to remain unidentified in the interest of encouraging victims to testify, thereby facilitating the prosecution of offenders.⁸⁸ *Attorney General v Leveller Magazine* (1979)⁸⁹ was also a contempt case, concerning publication of the name of “Colonel B”, a prosecution witness in an official secrets trial. The trial judge ordered, without protest from the defence, that Colonel B’s name be written down and shown to the court, the defendants and their counsel. The witness then testified in open court. The House of Lords in the ensuing contempt case declared that the concealment of witnesses’ identities was an extension of the court’s inherent power to hear cases *in camera* when required for the due administration of justice.

These two cases were used by the trial judge and by the Court of Appeal in Northern Ireland to justify the extensive use of anonymous witnesses in *R v Murphy and Maguire* (1989),⁹⁰ the first recorded case in which the

⁸⁶ *R v Legal Aid Board, ex p Kaim Todner (a firm)* [1998] 3 All ER 541, Lord Woolf MR at 548–49.

⁸⁷ [1974] 3 WLR 801.

⁸⁸ See Widgery LCJ, *ibid*, at 804.

⁸⁹ [1979] 2 WLR 247.

⁹⁰ Crown Court of Northern Ireland, April 1989, unreported; Court of Appeal in Northern Ireland [1990] NI 306.

identities of witnesses were withheld *from the defence*.⁹¹ This was the first of several trials arising from the murders of Corporals Wood and Howes in West Belfast in March 1988 during the funeral procession of Kevin Brady (killed at the funeral of the “Gibraltar 3” by Loyalist Michael Stone). A considerable amount of photographic and film material, obtained primarily from journalists reporting Brady’s funeral, was adduced by the prosecution in *Murphy and Maguire*. Twenty-seven media witnesses were called, essentially to confirm the material’s authenticity and to state the locations from which the images were filmed. In response to a prosecution request, the trial judge, Hutton LCJ, ruled that these witnesses were not to be identified even to the defence; that they would be screened from the view of the defendants, the public and the press; and that the media was not to disclose the name of any witness whose identity it discovered.

Murphy and Maguire is an important authority in the use of anonymous witnesses. But the basis on which Hutton LCJ made his ruling, the *Socialist Worker*⁹² and *Leveller*⁹³ cases, is flawed. In the *Socialist Worker* case, the witnesses’ names were concealed only from the press and the public, so the defence’s ability to cross-examine the witnesses was not limited by anonymity. Similarly, in the *Leveller* case the defence knew the identity of Colonel B. Indeed, at trial the magistrates refused the prosecution’s application for the identity of a different witness to be concealed from the defence. Lord Diplock, in the contempt hearing, stated that the magistrates would not have had the power to make the order sought by the prosecution. These cases, then, provide authority merely for shielding witnesses from *publicity*. However, the appeal court in *Murphy and Maguire* was unconcerned about the lack of precedent for withholding witnesses’ identity from the accused, stating that the absence of authority “indicates, in our view, nothing more than that the exceptions to the general rule [of open justice] are inevitably various and will continue to be so, for fresh circumstances will create new exceptions.”⁹⁴

The *Murphy and Maguire* case illustrates the adage that hard cases make bad law. It marks the incursion of the anonymous witness into “ordinary” criminal trials, a development which was eased by the exceptional circumstances of the case. It is precisely in such highly emotive and politicised trials that the observation of fundamental principles of criminal justice is most needed, yet is most easily dispensed with. This was truly a crime against the state: the murder of two British corporals, at a Republican funeral, by members of the IRA. Its impact was heightened as images of the beatings and killings of the two soldiers were transmitted around the world. It was perhaps predictable that Lord Chief Justice Hutton would accede to the prosecution’s request for the media witnesses to testify anonymously, especially as the crown submitted its intention to call the same witnesses in

⁹¹ Although there is authority for the proposition that a defendant can be removed from the courtroom if a witness is likely to be intimidated (*R v Smellie* (1919) 14 Cr App R 128, CCA) and the use of screens, particularly in cases of child sexual abuse, is now well-established, there was no precedent at the time of *Murphy and Maguire*’s trial for the concealment of witnesses’ identities.

⁹² *Op cit* n 87.

⁹³ *Op cit* n 89.

⁹⁴ [1990] NI 360; at pp 89-90 of transcript.

dozens of subsequent, related trials. The trial procedures in *Murphy and Maguire* were abnormal, even in the context of the criminal justice system of Northern Ireland, and the use of anonymous witnesses was unsupported by precedent. Yet the case has been used in England and Wales to establish precedent for the granting of anonymity in routine trials.

The first British case in which civilian prosecution witnesses testified anonymously was *R v Brindle and Brindle* (1992),⁹⁵ the trial of two brothers for murder. The crown sought anonymity for three material witnesses on the basis of their fear of providing testimony. The trial judge, Denison J, acceded to the crown's request, commenting that,

"If there is a real danger that a witness will not give. . . [material] evidence because of a genuine fear of the consequences if his identity becomes known, *then the witness must be given such protection as the Court is able to provide*. I use the word 'genuine fear' as opposed to 'justified fear' because it seems to me that *it is the state of mind of the witness which is vital*. If he is afraid, and if that fear is genuine, then it is not conclusive that no direct threat has been made to him. It seems to me it is sufficient if he genuinely fears retribution and that fear may stem from what some might call rumour or gossip, but which others would call local knowledge."⁹⁶

To regard a witness' fear as decisive to a claim for anonymity accords inadequate weight to the right to a fair administration of justice, as the European Court of Human Rights recognised three years prior to the *Brindle* trial.⁹⁷ Moreover, the Court would not countenance the granting of anonymity on the subjective basis proffered by judge Denison.⁹⁸ His judgment continued to reflect a superficial consideration of the defendant's legal rights:

"I recognise that the interests of the defence require that witnesses be seen by the defendant, and if they are not, then there must be. . . some inhibition on the full and proper presentation of the defence. [But] if the wider interests of justice make it necessary for anonymity. . . then the interests of the defence must be subordinated to those wider interests; in particular, that material evidence which is available should be given at the trial."

In accordance with the judge's ruling, all identifying information relating to three material witnesses was withheld from the defendants, their counsel and from the press and public. In addition, the witnesses were screened from the view of the defendants, the press and the public gallery. Although the order related to three specific witnesses, when the trial commenced it became

⁹⁵ Central Criminal Court, 31 March 1992, unreported; emphases added; transcript extracts are reproduced by permission of Newgate Reporters Ltd.

⁹⁶ Judge Denison added an inconsistent qualification: "Of course a wholly irrational fear without any foundation could not be regarded as genuine and would not suffice."

⁹⁷ See *Kostovski v The Netherlands* (1989) 12 EHRR 434, para 44.

⁹⁸ See *Doorson v The Netherlands* (1996) 22 EHRR 330, para 71 at 358 – 59; *Van Mechelen* (1998) 25 EHRR, para 61.

apparent that all civilian witnesses were to be afforded anonymity. When the defence objected, the witnesses were questioned in the absence of the jury to ascertain their reasons for seeking protection; two admitted they had no need of anonymity. After this examination of the witnesses, all but two were granted anonymity. Denison J observed that his ruling was preferable to the admission of written witness statements, without testimony, under section 23 of the Criminal Justice Act 1988. However, although the Act provides for this arrangement where a witness is too afraid to testify, it also allows for the adduction of evidence relevant to the witness' credibility in order to compensate for the inability to conduct cross-examination.⁹⁹ Whilst principled objections may be made to the admission of section 23 statements, the judge erred in his view that the anonymity order was less restrictive, given the defence's inability to examine credibility where the witness' identity is unknown. Moreover, the *Brindle* case appears to be the first United Kingdom case in which *material* witnesses were granted anonymity, representing an extension of the ruling in *Murphy and Maguire* (heavily relied on by the crown in *Brindle*). Following the acquittal of the Brindle brothers, journalist Caroline Godwin (aided by two barristers who were so concerned that they acted on a *pro bono* basis) unsuccessfully sought judicial review of Denison J's decision. Richard Palmer, writing in *The Sunday Times*, observed that, "The predicament is becoming increasingly common for court reporters in Britain. Secret justice is being handed out in a growing number of cases, often for the flimsiest of reasons."¹⁰⁰

In the same year as the *Brindle* case, the crown sought witness anonymity in a committal for trial of six defendants charged with violent disorder. Watford Magistrates' Court ordered that eleven witnesses, some of them alleged victims, were to be afforded anonymity through the ascription of colours as pseudonyms, screening from the defendants and testimony via voice-distortion equipment. The committal proceedings were stayed pending the defendants' application for judicial review of the ruling (*R v Watford Magistrates' Court, ex p Lenman*).¹⁰¹ In response to the applicants' assertion of a common law right of confrontation, Beldam LJ in the Divisional Court declared:

"It is now well established that there may be occasions upon which the interests of justice require that the identity of witnesses should be withheld. This will only rarely be done and on valid cause being shown. Equally, of course, it would be pointless to withhold the identity of the witnesses or the means by which they could be identified if at the same time the circumstances in which they gave their evidence were such that they could by other means, either because of their appearance, or because of the sound of their voices, easily be identified."¹⁰²

⁹⁹ Criminal Justice Act 1988 sch. 2(1).

¹⁰⁰ *The Sunday Times*, 15 August 1993.

¹⁰¹ [1993] Crim LR 388.

¹⁰² *R v Watford Magistrates' Court, ex p Lenman and Others*, QBD, 7 May 1992; transcript.

Despite Beldam LJ's statement that anonymity will be dependent on "valid cause" and on "substantial grounds being shown", his judgment accepts that an assertion of fear on the part of a witness will suffice. Indeed, his Lordship observed that the Divisional Court would not interfere with an anonymity ruling unless it was *Wednesbury* unreasonable.¹⁰³ Furthermore, the Court failed to illustrate its assertion that the power to grant witness anonymity was well established. The only case cited in the Court's judgment was that of *R v XYZ* (1990),¹⁰⁴ an authority for the use of screens which did not involve withholding witnesses' *identities*.

In common with the other cases involving anonymous witnesses, the Divisional Court in *ex p Lenman* failed to address the accused's specific rights. The applicants contended that the effect of the magistrate's order would be to deprive them of a line of cross-examination: without being able to see a witness (and thus ascertain or challenge his precise location and his behaviour at the scene of the violent disorder), they would be unable to instruct counsel to ask pertinent questions. Beldam LJ responded that since a witness would have given his location in his police statement, the applicants would not be so hampered. This riposte, however, ignores the wider restriction on challenging the witness' credibility beyond the circumstances of the events in question. The judgment gives no sense of engagement with fundamental principles:

"If the rights of an accused and in particular his ability to prepare for and conduct his defence are. . . prejudiced, justice requires the court to balance the prejudice to him and the interests of justice. . . [I]t is difficult to see that there is any but a theoretical prejudice to the applicants in not knowing the identity of. . . [a] person or his precise physical characteristics. . . but there is clear evidence of fear before the magistrate on the part of the witnesses and an unwillingness on their part to give evidence unless protection was provided for them and their identity was not disclosed."¹⁰⁵

The public interest in securing witnesses' testimony may justify anonymity in exceptional cases. However, the state has a duty to observe, even facilitate, a defendant's legal rights; any restriction of these entitlements

¹⁰³ "It is difficult to think of a decision more dependent on the exercise of discretion than the Magistrate's decision in this case. It is plain that this court will not interfere with such a decision unless it can be shown to be so unreasonable that no magistrate, properly considering the question and properly directing himself, could have reached that conclusion." Whilst this would appear to limit a challenge to the strong sense of *Wednesbury* unreasonableness, it is not altogether clear whether his Lordship intended to exclude *Wednesbury* unreasonableness in its broader sense; if the former, then successful challenges are highly unlikely. For explanation, see P Leyland and T Woods, *Textbook on Administrative Law* (3rd ed, 1999; London: Blackstone Press), pp 225 – 227.

¹⁰⁴ 91 Cr App R 36. Screens were used in this case solely for the purpose of protecting child victims of sexual abuse from being exposed to the gaze of the defendants.

¹⁰⁵ *R v Watford Magistrates' Court, ex p Lenman and Others*, QBD, 7 May 1992; transcript; emphasis added.

must accord with principle and be accompanied by adequate procedural safeguards.

5. UK JURISPRUDENCE AND ARTICLE 6 OF THE ECHR

(a) Current UK jurisprudence

The leading authority in England and Wales on witness anonymity is the Court of Appeal's judgment in *R v Taylor* (1995).¹⁰⁶ The trial judge, Denison J (who also presided over the *Brindle* trial), acceded to a prosecution request for witness anonymity. "Miss A" gave material evidence for the prosecution in a trial for murder and perverting the course of justice. Her name and address were withheld from the defence and she was protected from sight of the defendants by a screen (though they could see her via a video camera). Despite an admission by Miss A in cross-examination that she was not as fearful of testifying as had been supposed, the judge refused to reconsider the anonymity order. Denison J explained the reason for the order thus:

"This is one of those cases where a balancing act has to be done. On the one hand you have got the requirement of open justice and the hindrance to cross-examination, although. . . in this case I do not think the hindrance is very great. On the other you have got a witness who is frightened. It seems to me. . . that the solution proposed by the Crown is the best compromise and, therefore, I will allow this girl to remain anonymous."¹⁰⁷

The Court of Appeal in *Taylor* observed that a defendant's right to see and to know the identity of his accusers should only be denied "in rare and exceptional circumstances." But the court made cursory reference to the cases of *Smith v Illinois*¹⁰⁸ and *S v Leepile*¹⁰⁹ on the basis that they indicated an absolute right of confrontation, and the further case of *R v Murphy and Maguire*¹¹⁰ which was curiously interpreted as concerning the use of screens, rather than anonymity. The appellate court in the instant case adopted the decision in *ex p Lenman*¹¹¹ and endorsed Denison J's anonymity ruling, declaring that it was "close to being a model exercise" of discretion.

Evans LJ, giving the judgment of the Court in *Taylor*, identified five considerations which may justify an anonymity order:¹¹² there are real grounds for fearing the consequences of testifying as a named witness; the evidence is so important that it would be unfair to deprive the prosecution of it; the prosecution has fully investigated the witness' credibility; no undue prejudice would be caused by allowing anonymous testimony; and the court can balance the need for anonymity against unfairness to the defendant.

¹⁰⁶ [1995] Crim LR 253.

¹⁰⁷ Cited by the Court of Appeal, *R v Taylor and Crabb*, 22 July 1994, transcript.

¹⁰⁸ 390 US 129 [1968].

¹⁰⁹ *Op cit* n 20.

¹¹⁰ *Op cit* n 90.

¹¹¹ *Op cit* n 101.

¹¹² The Court noted that the factors were applicable regardless of whether the application is made by the prosecution or the defence.

Whilst the requirement of “real grounds” for fearing revelation of the witness’ identity appears exacting, it has not raised the threshold at which a witness’ fear justifies anonymity. The Court made no reference to any requirement of evidence of a threat to the witness, neither did it mandate an *objective* assessment of the likelihood of danger. Indeed, Evans LJ, consistent with the approach of Denison J in *Brindle* and Beldam LJ in *ex p Lenman*, repeatedly implied that an *assertion* of fear would suffice. However, ECHR jurisprudence requires an assessment of the risk of threat before the use of anonymous witnesses can accord with the defendant’s Convention rights.¹¹³ The Court of Appeal in *Taylor* undermined the first prerequisite as a safeguard for the defendant in two further respects: if the witness is not herself afraid, but others fear for her, that will justify the granting of anonymity to the witness; regardless of who is afraid, the consequences feared need not relate to the witness (for instance, they could presumably extend to her family).

There is an obvious problem in relation to the second factor identified in *Taylor*, that the witness’ testimony is so important that it would be unfair to effectively deprive the prosecution of it by denying anonymity. As the Court itself recognised, the more important the evidence, the greater the prejudice to the defendant in permitting anonymity. Furthermore, judges are to consider whether the credibility of the witness is (or is likely to be) in issue or whether the question for the jury is simply one of the reliability and accuracy of the witness, in which case anonymity will more readily be granted. This fails to address the reality that notwithstanding its pre-trial disclosure obligations the prosecution alone is in a position to know whether the witness’ credibility is questionable. There is little comfort for the defence in the third factor outlined in *Taylor*: whilst the court must be satisfied that the prosecution has thoroughly investigated the witness’ credibility, the outcome of that inquiry must be disclosed to the defence “so far as is consistent with the anonymity sought.”

With regard to the penultimate *Taylor* consideration, that granting anonymity would cause no undue prejudice to the defendant, the Court failed to acknowledge the nature and extent of the ensuing disadvantage. “‘Undue’ is a necessary qualification because some prejudice is inevitable if the order in question is made, even if that prejudice is *only* the qualification placed on the right to confront the witness as one of the defendant’s accusers.”¹¹⁴ The actual prejudice engendered is the undermining of the presumption of innocence (because the defendant is insidiously identified as “dangerous” and, perhaps, “guilty”) and of the requirement of full proof (because a prejudiced jury might be seduced into convicting where the evidence has not eliminated all reasonable doubts). Yet in the Court of Appeal’s view this does not amount to “undue” prejudice. Indeed, it is a feature of the cases involving anonymous witnesses that the risk of prejudice to the accused is regarded as marginal (“theoretical”, even, in *ex p Lenman*) whilst the witness’ fear is referred to in concrete terms.

The final matter to which a judge must have regard when adjudicating an application for witness anonymity is also unsatisfactory. The fundamental

¹¹³ *Van Mechelen v The Netherlands*, *op cit* n 83.

¹¹⁴ *Op cit* n 106, transcript; emphasis added.

difficulty with balancing the need for witness protection against the (appearance of) unfairness to the accused is the way in which the need for protection is assessed, which remains problematic in light of the guidance from *Taylor* taken as a whole. Furthermore, the Court of Appeal noted that the balancing of protection against unfairness includes the *extent* of protection which might be afforded to witnesses but provided no guidance as to the appropriate levels of protection in particular circumstances.

(b) ECHR jurisprudence

The *Taylor* guidance was approved by the Government's interdepartmental working group on vulnerable and intimidated witnesses.¹¹⁵ However, current United Kingdom law does not satisfy the requirements of ECHR jurisprudence and might not, therefore, survive a challenge under the Human Rights Act 1998.

(i) Article 6 principles

Article 6 of the ECHR has been developed by the European Court of Human Rights to provide both general procedural principles and specific rules relating to the use of anonymous witnesses. At a general level, Article 6 requires all evidence to be adduced at a public hearing, with a view to

adversarial argument.¹¹⁶ There are exceptions to this principle,¹¹⁷ but in any event defence rights must not be *infringed*.¹¹⁸ Any measures *restricting* defence rights must be strictly necessary and the least restrictive procedure possible must be used.¹¹⁹ A further general principle is that the defendant must have a proper opportunity to challenge a prosecution witness.¹²⁰ This opportunity will be severely restricted where the defence is confronted with an anonymous witness, as the Court recognised in *Kostovski v The Netherlands* (1989):

“If the defence is unaware of the identity of the person it seeks to question, it may be deprived of the very particulars enabling it to demonstrate that he or she is prejudiced, hostile or unreliable. Testimony or other declarations inculcating an accused may well be designedly untruthful or simply erroneous and the defence will scarcely be able to bring this to light if it lacks the information permitting it to test the author's

¹¹⁵ *Op cit* n 3.

¹¹⁶ *Barbera, Messegue and Jabardo v Spain*, 1988, Series A no. 146; *Kostovski v The Netherlands op cit* n 97.

¹¹⁷ For instance, Article 6(1) permits the exclusion of the press or public from all or part of the trial “in the interest of morals, public order or national security. . . where the interests of juveniles or the protection of the private life of the parties so require, or to the extent strictly necessary. . . in special circumstances where publicity would prejudice the interests of justice.”

¹¹⁸ *Van Mechelen v The Netherlands, op cit* n 83.

¹¹⁹ *Ibid.*

¹²⁰ *Unterpertinger v Austria* (1986) 13 EHRR 175; *Kostovski v The Netherlands op cit* n 97; *Windisch* case, judgment of 27 September 1990, Series A no. 186; *Ludi v Switzerland op cit* n 5; *Doorson v The Netherlands op cit* n 5; *Van Mechelen v The Netherlands op cit* n 83.

reliability or cast doubt on his credibility. The dangers inherent in such a situation are obvious.”¹²¹

ECHR jurisprudence has been influenced by the fact that the cases examined have involved anonymous witnesses who testified neither at the trial nor at the appeal stages in the national courts.¹²² Whilst the Strasbourg judicial authorities expressed strong concern about the use of anonymous witnesses in *Kostovski v The Netherlands* (1989),¹²³ *Windisch v Austria* (1990)¹²⁴ and *Ludi v Switzerland* (1992),¹²⁵ what appears to have been decisive in these adverse judgments was the denial of the accused’s right of confrontation. This interpretation is supported by two failed challenges to the use of anonymous witnesses: *Baegen v The Netherlands* (1994)¹²⁶ and *Doorson v The Netherlands* (1996).¹²⁷ In *Baegen*, although an anonymous victim of rape did not testify, the defence had been able to challenge her credibility at the appeal stage (the accused’s counsel having discovered her identity).¹²⁸ In *Doorson*, although the anonymous witnesses did not testify in open court, the defence lawyer was permitted to question them (albeit in the absence of the accused) at appellate stage in a hearing before the investigating judge. Nevertheless, although the dominant theme of the jurisprudence is the absence of witnesses rather than their anonymity, it has been established that affording the defence the opportunity to question anonymous witnesses will not, of itself, satisfy the requirements of Article 6 of the Convention.

Turning to more specific principles, challenges to the use of anonymous witnesses have required the Commission and the Court to consider the scope of the accused’s right in Article 6(3)(d) “to examine or have examined witnesses against him.” In *Kostovski*, material prosecution evidence was supplied by two informants, neither of whom testified at trial, though their statements were admitted as evidence. One of the informants was interviewed, in the absence of the defendant and his lawyer, by the examining magistrate who did not know the informant’s identity. The other informant was interviewed only by the police. The defence was subsequently invited to submit written questions to the first informant (who, on the basis of anonymity, responded to only two of the fourteen questions) but had no opportunity to question the second. At trial, no questions about the informants’ reliability and sources were permitted where the answers might

¹²¹ *Op cit* n 97, para 42.

¹²² If the prosecution uses a statement given by a person who does not attend the trial, that person is considered a “witness” for the purposes of Article 6: *Windisch case op cit* n 120; *Ludi v Switzerland op cit* n 5; *Kostovski v The Netherlands op cit* n 97.

¹²³ *Op cit* n 97.

¹²⁴ *Op cit* n 120.

¹²⁵ *Op cit* n 5.

¹²⁶ Application 16696/90; Report of the Commission adopted on 20 October 1994.

¹²⁷ *Op cit* n 5.

¹²⁸ The Commission was also influenced by the defence’s failure to take opportunities to submit written questions to the anonymous witnesses, the accused’s refusal to provide police with samples of blood and saliva, the right to respect for privacy of victims of sexual offences and the fact that the conviction did not rest solely on the statements of the anonymous witnesses (though their statements were arguably decisive). The case was not considered by the European Court of Human Rights.

have revealed their identity. The European Court of Human Rights observed that although the defence had been able to question two examining magistrates, a police officer and, indirectly, one of the informants, the scope and nature of the permissible questioning was significantly limited by the decision to preserve anonymity. Furthermore, the trial and appellate courts had no opportunity to assess the informants' reliability independently through observation of their demeanour. In finding a violation of Article 6, the Court stated that witness intimidation, though a serious concern, could not justify the denial of a fair administration of justice. The Court held that where anonymous witnesses are used, the disadvantages to the defence must be counterbalanced by the procedures adopted by national judicial authorities. Moreover, the Court established that whilst reliance at investigative stage on anonymous informants is not necessarily incompatible with the Convention, a conviction can *never* be founded solely or to a decisive extent on the statements of anonymous witnesses.¹²⁹ The failure of national courts to observe these principles led to adverse judgements in *Windisch v Austria* (1990)¹³⁰ and *Ludi v Switzerland* (1992),¹³¹ in which the defence had been denied the right of confrontation with absent, anonymous witnesses whose identity was not disclosed to the trial and appellate courts.

(ii) Specific procedural requirements

A detailed examination of the procedure challenged in the leading cases on anonymous witnesses, *Doorson v The Netherlands*¹³² and *Van Mechelen v The Netherlands*,¹³³ demonstrates the inadequacy of the *Taylor* guidance provided by the United Kingdom Court of Appeal.¹³⁴ In *Doorson*, the defendant was convicted of drug dealing; the evidence of two named witnesses ("R" and "N") and two anonymous witnesses ("Y15" and "Y16") was decisive. All the witnesses were drug addicts who had identified the accused from police photographs; at no stage in the proceedings was there a direct confrontation between the defendant and R, Y15 and Y16. Prior to the trial the investigating magistrate, contrary to an undertaking, interviewed the anonymous witnesses in the absence of Doorson's lawyer. The Regional Court refused to call those witnesses but received their statements as evidence. The named witness R did not comply with the Court's summons;

¹²⁹ Confirmed in *Doorson v The Netherlands op cit n 5*; *Van Mechelen v The Netherlands op cit n 83*.

¹³⁰ *Op cit n 120*.

¹³¹ *Op cit n 5*.

¹³² (1996) 22 EHRR 330.

¹³³ (1998) 25 EHRR 657.

¹³⁴ Following the European Court's decision in *Kostovski*, the Netherlands Supreme Court provided procedural guidance on the use of anonymous witnesses ((1990) NJ, No 692). The guidance influenced the appellate stages of the *Doorson* and *Van Mechelen* cases. On many points, the Supreme Court's guidance was less rigorous than that established in *Taylor*, but it did require judicial assessment of the witness' reliability and of his/her reasons for desiring anonymity. The Supreme Court also adopted the European Court's ruling that a conviction should not be decisively based on the statements of anonymous witnesses and further required the trial court to explicitly state that the statements had been used with caution and circumspection. Subsequently, the Protection of Witnesses Act 1993 amended the Code of Criminal Procedure (section 226a) to regulate the use of anonymous witnesses.

N retracted his police statement, explaining that he had lied in order to secure the return of his confiscated drugs. On appeal against conviction, the Amsterdam Court of Appeal requested that the investigating judge re-examine Y15 and Y16 in order to verify the need for anonymity and to provide Doorson's lawyer with the opportunity to question the witnesses. The Court of Appeal accepted the judge's report that the anonymous witnesses (whose identities were known to her) had firm grounds for anonymity: Y15 had been threatened by drug dealers in the past and Y16 had suffered violent reprisal; both stated that they were afraid of the defendant. Nine months later at the request of the Court of Appeal, the investigating judge, without further examination of the anonymous witnesses, provided a limited endorsement of their reliability. In relation to the named witnesses, R absconded before he could be heard and N repeated the retraction of his police statement. The Court heard a further named witness, a police officer who testified that although he had never known the defendant to be violent or to issue threats, such future behaviour could not be discounted as drug dealers often exacted reprisals against those who incriminated them.

In a majority decision, the Commission found that there had been no violation of Article 6 because the investigating judge knew the identity of the anonymous witnesses (who had a well-founded desire to remain unidentified) and because the defence lawyer had been able to question them at the hearing before the judge. The European Court reached the same decision, noting that although the anonymity of Y15 and Y16 had "presented the defence with difficulties which criminal proceedings should not normally involve,"¹³⁵ the procedures employed by the national courts afforded the defence sufficient compensation. The Court found additionally that although fear of reprisal was a relevant reason for witness anonymity, it was not necessarily a sufficient one. In the present case, anonymity was justified on the basis of police knowledge that drug dealers were frequently violent to those who had testified against them and by reason of the fears expressed by the unidentified witnesses. The Court was satisfied that the Amsterdam Court of Appeal had treated the anonymous statements with extreme care and that it had not based its conviction to a decisive extent on that evidence. Both these observations may, however, be criticised. In *Kostovski*, the European Court declared that a cautious evaluation of statements by anonymous witnesses "can scarcely be regarded as a proper substitute for direct observation."¹³⁶ In relation to the basis of the conviction in *Doorson*, the Amsterdam Court stated that its decision was founded mainly on the evidence of the four witnesses; yet R was heard only by the police and N consistently testified that he had lied in his police statement. In these circumstances, it seems inescapable that the statements of the anonymous witnesses were decisive.

In *Van Mechelen v The Netherlands*, the defendants were convicted of attempted homicide and robbery with the threat of violence. The only identification evidence consisted of statements by eleven anonymous police officers. The officers were not interviewed by an investigating judge prior to the trial and they did not testify before the Regional Court, which remained

¹³⁵ (1996) 22 EHRR 330, para 72 at 359.

¹³⁶ *Kostovski v The Netherlands* (1989) 12 EHRR 434, para 43.

unaware of their identity. The Court of Appeal referred the case back to the investigating judge to clarify the basis for anonymity and to examine the witnesses in the presence of the defence. These hearings took place in two separate rooms connected by a sound link. The investigating judge, each witness and the registrar were present in one room, while the defendants, their lawyers and the Advocate General were in the other. Most of the police officers explained that they wished to remain anonymous to protect their families and friends, with some giving the additional reason of preserving their operational value as undercover officers. The defence was given the opportunity to question the witnesses at length. When the Court of Appeal hearing resumed (at which the defendants were convicted), the reasons for maintaining anonymity were explained:

“Of the arguments for continuing the anonymity of the witnesses, the Court of Appeal considers particularly persuasive the personal safety of these witnesses and their families, and it makes no difference that these witnesses have not yet been threatened. . . [T]he present case concerns extremely serious crimes, the [attempted murder] having been committed so as to evade recognition and arrest by the police, the perpetrators having been prepared to sacrifice a number of human lives. In these circumstances, the risk run by the witnesses identified only by a number and their families if their anonymity is lifted or insufficiently guaranteed is decisive. In so far as anonymous witnesses have refused to answer questions this was done in order not to disclose methods of investigation or to maintain the anonymity of other investigating officers involved in the case.”¹³⁷

Nevertheless, the European Court found that the use of anonymous witnesses was incompatible with Article 6. The Court distinguished the case from that of *Doorson*: in the latter the foundation for anonymity was more clearly established, the anonymous witnesses were heard in the presence of defence counsel and there was identification evidence from named witnesses. The tenor of the *Van Mechelen* judgment, however, conveys the impression that a similarly constituted Court would have found a violation in *Doorson*. Indeed, in light of the existing ECHR jurisprudence and of the arguably greater restrictions on the defence, *Doorson* was a surprising decision.

The *Van Mechelen* judgment refined and supplemented the existing Convention jurisprudence on anonymous witnesses in four main ways. First, although it had been established that it is incumbent on national courts to consider less restrictive forms of witness protection, the Court made it clear that this obligation will not be discharged simply by declaration. A related requirement of Article 6 is that prior to granting anonymity, courts must conduct an informed assessment of the risk of reprisals against witnesses seeking such protection. Specifically, the court should assess the ability of the defendant to exact reprisals or to secure others to do so on his behalf. The European Court requires national courts to demonstrate that this inquiry has been conducted. The second main development in *Van Mechelen* is that by virtue of their connection with the State, police officers should be given

¹³⁷ Cited by the European Court of Human Rights (1998) 25 EHRR 657, para 26.

anonymity only in exceptional circumstances. Furthermore, the desire to preserve the operational usefulness of undercover officers will not, of itself, justify anonymity: there must also be an established threat to the safety of the officers or their families. Thirdly, the defence must be able to question anonymous witnesses *in their presence*. Finally, the Court established for the first time that the *defence* must be able to assess the reliability of anonymous witnesses through observation of their demeanour. This is a significant development: previous cases had emphasised the importance of the *court* being able to observe demeanour. The current position is that a judicial assessment of the witnesses' reliability and credibility will not suffice. Even if a judge, having ascertained the witnesses' identity, directly questions them about their reasons for seeking anonymity and provides a detailed account of his findings as to their reliability and credibility, this will not represent sufficient compensation for the disadvantages caused to the defence. The Court regarded the entitlement of the defence to assess reliability as distinct from the right to question witnesses in their presence. It follows that the defence must be permitted to *see* the anonymous witness; the fact that the defence in *Van Mechelen* was able to hear the witnesses was not, in the Court's view, an adequate indicator of reliability. This decision implicitly acknowledges the importance of visual clues to the development of cross-examination strategy. Furthermore, it will allow the defence to challenge the assessment of a witness' reliability presented by the prosecution (and the judge).

(iii) *Compatibility of domestic law with ECHR jurisprudence*

Domestic law regulating the use of anonymous witnesses satisfies neither the general tenets of Article 6 of the Convention, nor the specific principles developed by the European Court of Human Rights from the Article 6 (3) right of confrontation. According to ECHR jurisprudence, national courts must conduct a thorough and objective examination of the existence and level of any risk of harm to a prosecution witness before granting anonymity. Furthermore, anonymity should only be resorted to if other forms of witness protection would not suffice. The Court of Appeal in *Taylor*,¹³⁸ however, indicated that a mere assertion of fear, without any evidence of a threat to the witness, would justify concealment of identity. Although the Court noted that judges could consider protective measures short of anonymity, it did not direct them to do so. There has been no appreciation in domestic courts of the particular concern surrounding the granting of anonymity to law enforcement officers.

It is a well-established Convention principle that where anonymous witnesses are used, the consequent limitations on the defence must be counterbalanced by procedural measures. These measures must provide the defence with the opportunity to question the witnesses in their presence and to assess reliability through direct observation of demeanour. United Kingdom law, however, permits the adduction of witness statements where a witness does not attend the trial.¹³⁹ Indeed, the House of Lords has endorsed the admission at coroners' inquests of statements by anonymous soldiers

¹³⁸ *Op cit* n 106.

¹³⁹ Criminal Justice Act 1988, s 23.

exempted from attending the hearings even though they were suspected of causing the deaths.¹⁴⁰ In relation to observation of demeanour, United Kingdom jurisprudence accords with Article 6 insofar as anonymous witnesses have not been screened from defence lawyers during testimony. Nevertheless, the jury's ability to assess credibility remains problematic by virtue of the prosecution's control of witness information, the enhanced status of anonymous witnesses and, in terms of physical indicators of veracity and accuracy, the use of screens and voice-distortion equipment.

The European Court of Human Rights has emphasised the importance of regarding anonymous statements with extreme caution. As Jane Liddy has observed,¹⁴¹ permitting unidentified police officers to refuse to answer questions on the basis that investigation methods or other officers would otherwise be revealed does not exhibit caution. Operational secrecy is a privilege routinely extended to law enforcement witnesses, but when combined with personal anonymity (as in the *Jack* case),¹⁴² the dangers in relying on the testimony of such witnesses are increased. Indeed, to base a conviction solely or to a decisive extent on anonymous statements contravenes Article 6. Yet the Court of Appeal in *Taylor* stated that the prospect of anonymity increases commensurately with the importance of the prosecution witness' testimony, raising the possibility of a direct contravention of the Convention.

6. CONCLUSION

There are two principal difficulties with the United Kingdom courts' approach to witness anonymity: the conception of anonymity as impacting chiefly on open justice (itself viewed at an abstract level, rather than taking cognisance of the accused's procedural, legal rights) and the notion of "balance" in the criminal justice system. Of course, fairness in criminal procedure does not reside exclusively in the domain of defendant rights: it is trite to note that in order to sustain the public support vital to the legitimacy of the criminal justice system, the courts must give expression to the public interest in the successful prosecution of offenders. But the traditional view of criminal justice, that it involves a balancing act between the conflicting objectives of the effective implementation of the criminal law and the protection of the defendant, leads inexorably to the conclusion that safeguards instituted to protect those accused of criminal conduct represent a hindrance to the aim of crime suppression.¹⁴³ Yet due process is not inconsistent with the notion of crime suppression: as a normative model, it prescribes the *procedure* to be employed in the prosecution of offenders. Although the due process model is commonly seen to imply a reduction in the efficiency of the criminal process, this view is predicated on the notion

¹⁴⁰ *Devine v Attorney General for Northern Ireland, Breslin v Attorney General for Northern Ireland* [1992] 1 All ER 609.

¹⁴¹ *Van Mechelen and others v The Netherlands*, Application Nos 21363/93; 21364/93; 21427/93; 22056/93, Report of the European Commission of Human Rights, 27 February 1996. The Commission concluded (20:8) that there was no violation of Article 6; Jane Liddy provided a dissenting opinion.

¹⁴² *Op cit* n 79.

¹⁴³ See further, J Jackson, "Getting Criminal Justice out of Balance" in J Morison and S Livingstone (eds), *Law, Society and Change* (1991), pp 114 – 133.

that fact-finding reliability is of secondary importance as a value. But public confidence is not secured simply by high rates of prosecution and conviction, as the reaction to publicised miscarriages of justice has shown; adherence to due process is essential to the very legitimacy of the criminal justice system. The choice need not be viewed as one of giving pre-eminence either to crime control values or to those reflecting due process. Different values can be, and to an extent are, accommodated in our criminal justice system. If witness anonymity is necessary in exceptional cases, this may be afforded provided the limitations consequent on the defence are to some extent compensated for and defence rights are not infringed.

The courts' current response to the problem of witness intimidation violates fundamental principles of criminal justice and contravenes international human rights law. There should be a rigorous judicial approach to claims for witness anonymity, coupled with the implementation of effective safeguards for the defence. Although the judiciary has thus far failed to appreciate the nature and extent of the encroachment of anonymity on defence rights, it is to be hoped that the Human Rights Act 1998 will encourage judicial activism in the safeguarding of the legal right to a fair trial.