Northern Ireland Legal Quarterly

Volume 62 Number 1

EDITOR
PROFESSOR SALLY WHEELER

Queen's University Belfast
School of Law

Contents

Erratum

In *NILQ* 61(4) (winter 2010), in "Multilateral governance of financial markets: the case of sovereign wealth funds" by George Gilligan. At p. 400 Gilligan quotes directly from one of the US interview respondents. The latter stated that: "In 2007, FINRA [Financial Industry Regulatory Authority] slightly increased reporting requirements around sovereign investment . . ." In fact, as correctly stated in n. 49, it was the Committee on Foreign Investment in the United States (CFIUS) that issued the relevant notice. FINRA does not have regulatory responsibility for sovereign investment in the US.

NILQ 62(1): 1–32

Law and the neoliberal vision: financial property, pension privatisation and the ownership society

PADDY IRELAND

Kent University

1 Introduction

When the financial crisis first broke, there was much speculation about whether it would prove a historic watershed which marked the end of the neoliberal era. It is, of course, still early days but there are, as yet, few signs that it will. On the contrary, the great majority of policymakers clearly hope that by supplementing the unprecedented fiscal and monetary stimuli which have already been administered with a dose of spending cuts, mandated austerity and "regulatory reform", things will return (more or less) to normal, enabling a more sober and sustainable version of the old regime to emerge.[1] The result is that at the moment the remnants of social democracy seem more likely to fall victim to the crisis than neoliberalism. Even within the business community, however, there are doubts about how far it will be possible to resume business as usual. Richard Lambert, outgoing Director General of the Confederation of British Industry, for example, recently suggested that "the public and political response to what's happened will . . . have troubling consequences", criticising everything from short-termism, executive pay and business culture to shareholder value, hostile takeovers and foreign shareholders. Endorsing Lambert's comments, John Plender of the *Financial Times* argued that "something radical ha[d] to be done" if business was "not to suffer from a permanent legitimacy deficit", while Roger Bootle of Capital Economics thought there was "a real crisis of capitalism about all this".[2] These hints that serious problems might be imminent have been echoed on the left. With the "conventional mantras" about the ability of "free markets and free trade, private property and personal responsibility, and low taxes and minimalist state intervention in

1 The remarkable degree of establishment consensus about this has been reflected in the quiet but growing convergence around a regime of what Susan Watkins has called "regulatory liberalism" in which the main components of the old regime – unfettered capital movements, free trade, "small" states, and private ownership (not least of banks) – remain firmly intact: S Watkins, "Shifting sands" (Jan–Feb 2010) 61 *New Left Review* 5, p. 13. She describes the crisis as having been "eerily non-agonistic", p. 20.

2 R Lambert, "Does business have a role as a 'force for good' in creating a more sustainable economic, social and environmental future?", RSA/Sky Sustainable Business Lecture Series, 30 March 2010; J Plender, "To avoid a backlash, executives must act on pay", *Financial Times,* 2 April 2010; R Bootle, "Executive pay: what the experts say", *The Guardian*, 31 March 2010, www.guardian.co.uk/business/2010/mar31/executive-pay-what-the-experts-say. This nervousness perhaps explains the rather hysterical and paranoid business reaction to Vince Cable's party conference speech in 2010.

social provision" to create a world of plenty "sound[ing] increasingly hollow", David Harvey has argued, "a crisis of legitimacy looms".[3]

Certainly, social unrest is growing, both in the UK and elsewhere, and the public mood is likely to darken further as public spending cuts and the "age of austerity" begin to bite.[4] Equally importantly, perhaps, it is already clear that the neoliberal vision of a world of inclusive "ownership societies" – populated by self-reliant "worker capitalists" who own their own homes and, through pension funds, modest amounts of financial property – is unravelling. This is potentially very significant for, in places like the UK, this vision has not only animated and influenced much recent policy-making, including (as we shall see) the growing focus on investor protection, but provided an important source of meaning, motivation and hope in people's lives and helped to legitimate many of the changes which have occurred in recent decades. If John Prescott's claim that "we're all middle class now" seemed a tad premature in 1997, in the decade following, more and more people undoubtedly became "petty bourgeois . . . by imaginative orientation if not by economic fact".[5] Will these imaginings survive the economic facts which are likely to emerge in the coming years and, if they don't, what will the consequences be?

Using pension privatisation as a vehicle, this article examines the impact of the financial crisis on the neoliberal vision. It argues that the vision is underlain by an economically determinist account of economic and social development which has had a major impact not only on economic and social policy but on legal policy and thinking about law. This economic determinism, it suggests, is rooted in mistaken ideas about the nature of markets and property, and financial property in particular; ideas which underpin equally misguided conceptions of self-reliance which deny our growing interdependence. The article moves on to explore the nature of neoliberalism and the gulf between neoliberal ideology and neoliberal practice, before concluding by arguing that the crisis represents an opportunity to re-enrich legal scholarship by cleansing it of the debilitating economic determinism with which so much of it has become infected and engaging in a radical rethink of what regulatory reform should entail.

2 The rise of pension privatisation

Since the 1980s there has been a concerted move away from traditional, pay-as-you-go (PAYG), state-administered pension systems in which tax revenues are used to pay pension benefits, towards privatised systems in which privately managed pension funds invest savings made by individuals during their working lives in financial property – in intangible rights to receive future revenues, such as government stock, corporate shares and corporate bonds. The idea is that by the time of their retirement, the old will directly or indirectly own (or have temporary, time-limited, beneficial claims over) sufficient revenue rights to live off or to fund the purchase of an annuity. They will, in other words, be able to become pensioner *rentiers*.

There are important differences between the two systems. Firstly, whereas in traditional public, state-administered systems benefits are defined in advance, in privatised systems, as "defined contribution" schemes replace "defined benefit" schemes, benefits increasingly depend on the returns accruing to the financial property held in individual accounts. Secondly, in traditional PAYG systems, risk is pooled and collectivised, with the result that

3 D Harvey, *The Enigma of Capital* (London: Verso 2010), p. 217.

4 At the time of writing, there have already been outbursts of public anger in places such as Greece, Ireland, Spain and France, as well as in the UK, where university tuition fees have generated furious protests.

5 R Unger, *The Left Alternative* (London: Verso 2002), p. 47. Shortly before the 1997 election, John Prescott famously announced that "we're all middle class now".

PAYG systems tend to be redistributive between both classes and genders, and within and between generations. By contrast, in privatised systems, where pension benefits are tied to individual contributions, there is little or no redistribution within or between classes or genders; risk and reward are individualised. Unlike the financing of traditional PAYG systems, however, the financing of private pensions is said to be "fully funded", meaning that the fund from which the pensions are drawn is on hand when retirement commences, derived as it is from earlier saving. It is therefore suggested that in privatised systems pensioners do not impose a burden on the younger generation but provide *for themselves* in retirement using the returns accruing to (and/or capital value of) their financial property.[6]

The move towards pension privatisation began in Chile under Pinochet in the late 1970s, led by the then Minister of Labour and Social Security, Jose Pinera.[7] The existing public PAYG system was replaced by a fully funded privatised system in which the savings of workers are invested in financial property held in personal retirement accounts administered by private, profit-oriented enterprises. Under the scheme, existing employees are compelled to invest 10 per cent of their wages, with the option of investing up to another 10 per cent voluntarily. They are allowed to choose between a range of investment institutions offering different mixes of bonds, equities, home and foreign investments, giving them some say in how (riskily) their funds are invested. Retirement age is a matter for the worker, the size of whose accumulated fund determines his or her benefits. There is a state top-up to the individual savings accounts of those who have worked for at least 20 years but failed to accumulate sufficient capital to buy a minimum annuity; a low-level welfare pension, unrelated to years worked, is available to those in extreme poverty.[8]

The reforms implemented in Chile paved the way for pension privatisation elsewhere. In the UK, for example, the Thatcher government began to promote pension privatisation in the 1980s, increasing the state pension in line with prices rather than gross earnings, causing its relative value to fall significantly.[9] At the same time, as previously state-owned industries were privatised, the government began to encourage people to engage in direct or indirect ownership of financial property, introducing a variety of tax-favoured forms of saving and investment, a process continued by New Labour.[10] Crucially, the 1980s also saw various international agencies start to advocate shifting the management of social risk away from the state towards the individual. The International Monetary Fund (IMF) announced that it was no longer realistic to assume that the real value of state pensions could be maintained,[11] while the Organisation for Economic Co-operation and Development (OECD) began to dispense advice on how pension provision might be incrementally

6 See M Orenstein, *Privatizing Pensions* (Princeton: Princeton University Press 2008), pp. 179–93.

7 Pinera, a Chicago-trained economist is the brother of the current Chilean President (and dollar billionaire), Sebastian Pinera.

8 See A F Yazigi (ed.), *The Chilean Pension System* (Santiago: SAFP 2003). The contributions made by workers are tax deductible and the returns on the investments untaxed until withdrawal on retirement. On retirement, workers can choose to draw on their accumulated fund each month or use it to purchase an annuity from a life insurance company.

9 The state pension fell in value from 20% to less than 15% of average earnings between the late 1970s and 2001: O Attansio and S Rohwedder, "Pension wealth and household saving: evidence from pension reforms in the UK", *IFS Working Paper* WP01/21 (2001), p. 5. The coalition government recently announced that it intends to restore the earnings link for the basic state pension from April 2011, but also intends to raise the retirement age.

10 See J Banks and S Tanner, *Household Saving in the UK* (London: IFS 1999), pp. 89–90; P Saunders and C Harris, Privatization and Popular Capitalism (Buckingham: Open University Press 1994); P Ireland, "Shareholder primacy and the distribution of wealth" (2005) 68 *Modern Law Review* 49.

11 See D Henwood (December 1994) *Left Business Observer*, issue 67. The IMF was, by then, using loan conditionality to impose "structural reforms" on indebted countries.

privatised.[12] The seminal moment came in 1994 with the publication by the World Bank of *Averting the Old Age Crisis: Policies to protect the old and promote growth*, a policy report which proposed a multi-pillar approach to pension provision. Mandatory, defined contribution savings schemes in which workers' savings are invested by privately managed financial institutions provide the main pillar. Existing defined benefit, public pensions provide the second pillar, though offering only a safety net. Voluntary savings and annuity schemes provide a third pillar for those with the means and inclination to save more.[13] Promoted by the World Bank and other international agencies, particularly in South America and post-Communist Eastern Europe, these ideas steadily spread among policymakers and academics during the course of the 1990s.[14] Although the campaign met with opposition (led by Joseph Stiglitz) from within the World Bank and this had some impact on the advice offered, it was "mostly rebuffed by the pension reform community within the Bank".[15]

3 The economic case for pension privatisation

Neoliberalism and the rationality of the market

Why has the idea of privatising pensions attracted so much support? As the title of the World Bank report suggests, many believe that pension privatisation is the only way to defuse the "demographic time bomb" confronting state systems. Indeed, by the time it appeared, greater longevity was already being portrayed as a threat rather than as a cause for celebration. The report also, however, put a positive case for pension privatisation rooted in a specifically neoliberal vision of economic, social and individual development.

This case was composed of two strands, one economic and one ethical. The economic strand, which was highlighted by the emphasis on the "and" in the title of the World Bank report, centres on the claim that traditional PAYG pension systems have deleterious macro-economic effects, acting as a drag on growth, demanding higher rates of tax, inhibiting savings and investment, adding to labour costs and encouraging capital flight. By contrast, privatised systems are said to foster growth by encouraging saving and investment, reducing the cost of capital to business, stimulating the development of capital markets and promoting the efficient allocation of resources. If invested in the private sector, argues Estelle James, leader of the research team that produced the report, mandatory pension funds are likely eventually "to dominate the ownership of financial assets that represent claims on the economy's real assets" and this, she suggests, will tighten the stock market disciplines to which corporate managers are subject. Worker–investors, interested only in

12 See R Holzmann, *Reforming Public Pensions* (Washington: OECD 1988), arguing that "what was only a couple of decades ago considered as a central achievement of the welfare state is now being evaluated differently".

13 World Bank, *Averting the Old Age Crisis: Policies to protect the old and promote growth* (Oxford: World Bank/OUP 1994). It is not insignificant that it was only after the collapse of Eastern European Communism that the assault on public pensions began in earnest. Just as the rise of public pension provision was prompted in part by the desire to reduce support for radical working-class social movements, the growing weakness of organised labour and elimination of the communist threat smoothed the way for its erosion.

14 See R Holzmann, "Reforming old-age pension systems in central and eastern European countries in transition" (December 1993) 7 *Journal of Economics*, Supplement 1, 191–218; and Orenstein, *Privatizing Pensions*, n. 6 above, ch. 4. When New Labour came to power in 1997, Jose Pinera, by now one of pension privatisation's leading international advocates, was invited to meet the Select Committee on Social Security by Frank Field, an admirer of the emphasis private pensions put on personal responsibility: See B Jamieson, "UK's Frank Field invites Jose Pinera", *Sunday Telegraph*, 11 May 1997.

15 Orenstein, *Privatizing Pensions*, n. 6 above, p. 81. For a discussion of the internal World Bank debates, see ibid. pp. 80–4. See also P Orszag and J Stiglitz, "Rethinking pension reform: ten myths about social security systems" in R Holzmann and J Stiglitz (eds), *New Ideas about Old Age Security: Towards sustainable pension systems in the 21st century* (Washington DC: World Bank 2001).

maximising returns and minimising risk, will, through their institutional representatives, be able to monitor corporate performance and insist that efficiency and "shareholder value" are maximised.[16]

This economic case is rooted in the neoliberal theories of economic and social development that rose to prominence in the 1980s and which have exerted enormous influence not only on policy-making but on the trajectory of legal thought and scholarship. The key supposition of these theories is that free markets know best; that private, contractual economic ordering and the "unregulated" forces of supply and demand serve to maximise not only individual freedom but efficiency, growth, wealth and welfare.[17] Particular and specific emphasis is laid on the creation of open financial markets which enable capital to move freely around the world – the so-called "efficient capital markets hypothesis". These theories are strident in their assertion of the existence of naturally functioning and naturally efficient markets.[18] Indeed, it is belief in the existence of a transhistorical, beneficent, market-based, economic rationality – a rationality which tends to be depicted as not merely politically neutral but apolitical – that underlies the concept of "the market" as an abstract, singular phenomenon. If the technical, legal, political and cultural barriers inhibiting the operation of this rationality were removed, the argument runs, it would operate for the benefit of all.

These theories are underlain by a belief in the (potential) autonomy of "the market" and "the economic" from the legal and political, a belief reflected in the idea that states (spatially separate political powers) "intervene in" and/or "regulate" markets (spatially separate economic spheres).[19] As such they are premised on a naturalisation of markets and property rights; on an assumption that they somehow pre-exist law and regulation. Lawyers, of course, are, or should be, aware that this is not the case; that property and markets are legal, political and social constructs – the products as well as the objects of regulation; and that, as a result, not only is the goal of "deregulation" absurd,[20] the dynamics and rationalities of particular markets are themselves inevitably political and legal products which vary according to the legal rights–obligations–regulatory structures that constitute them.[21] This is especially true of international financial markets which exist under the authority and by permission of the state, and are conducted on whatever terms states choose to allow.[22] Unaware of – or choosing to ignore – this, neoliberals purport to seek a

16 E James (1998) "New systems for old age security" (August 1998) 13(2) *World Bank Research Observer* 5–8, reprinted in S Ringen and P de Jong, *Fighting Poverty: Caring for parents, children, the elderly and health* (Farnham: Ashgate 1999); and World Bank, *Averting*, n. 13 above, pp. 208–16. Pension privatisation is thus seen as desirable in the developing as well as the developed world. Indeed, until recently, Chile was regularly held out as a role model whose privatisation programme had "deepened the nation's capital market . . . stimulated economic growth" and helped to generate the high growth rates achieved from the mid-1980s to mid-1990s: M Skousen, *EconoPower: How a generation of economists is changing the world* (Hoboken: Wiley & Sons 2008), ch. 5. The slowdown which followed was portrayed by the system's supporters as an aberration, a product of the government's "erroneous response" to the Asian crisis of 1997: J Pinera, "Retiring in Chile", *New York Times*, 1 December 2004.

17 See D Harvey, *A Brief History of Neoliberalism* (Oxford: OUP 2005). The Chicago wing of neoliberalism emphasises the efficiency benefits of markets; the Austrian wing their freedom-enhancing capacities.

18 They are said to be rooted in human nature and the natural propensity of self-interested human beings "to truck, barter and exchange" identified by Adam Smith: *Wealth of Nations* vol. 1, R H Campbell, A S Skinner and W B Todd (eds) (Oxford: OUP [1776] 1976), p. 25.

19 See E Meiksins Wood, *Democracy against Capitalism* (Cambridge: CUP 1995).

20 See D Campbell, "The end of Posnerian law and economics" (2010) 73 *Modern Law Review* 305, p. 326.

21 See P Ireland, "Property, private government and the myth of deregulation" in S Worthington (ed.), *Commercial Law and Commercial Practice* (Oxford: Hart 2003).

22 See S Strange, *Casino Capitalism* (Oxford: Blackwell 1986), p. 29.

"depoliticisation of economic life" and the establishment of market mechanisms in as many spheres of social activity as possible.[23]

This hostility towards (certain sorts of) political intervention in the market elides into a more general hostility towards the state and collectivism,[24] with states commonly being depicted as ineffective, predatory collections of self-seeking politicians and bureaucrats and as vehicles for special-interest groups. Indeed, as doubts grow about their powers and capacities in a "globalised" world, many now see states as "at best as an anachronism and at worst as an obstacle to human progress".[25] While this scepticism was conveniently forgotten when the crisis broke,[26] it is rapidly re-surfacing as governments seek to implement and justify spending cuts. Neoliberal hostility towards the state also finds expression in its uncomfortable and ambivalent relationship with democracy. Notwithstanding the emphasis they place on the importance of individual choice and consent, for neoliberals there are few more serious threats than democracy. Feeling unable openly to oppose it in the same way as their nineteenth-century counterparts,[27] however, contemporary neoliberals have tried instead to immunise the economy and the market from democratic control. Hence the growing use (associated with the shift from "government" to "governance") of "experts" and independent agencies,[28] and the emergence of a "new constitutionalism" which limits the power of states by granting quasi-constitutional protection to property interests.[29] It is not insignificant, perhaps, that the first country to engage in thoroughgoing pension privatisation, Chile, was a US-backed dictatorship, nor that one World Bank study concluded that one of the "most important . . . negative lessons" of pension reform (in Argentina) was that "the democratic process" tended to "dilute" reform programmes.[30]

Dictatorship of no alternatives

These ideas are illustrative of the neoliberal tendency to naturalise not only the market but a very particular model of capitalism. If the collapse of communism "seemed to confirm . . . that capitalism is the natural condition of humanity",[31] for many neoliberals the relative success of Anglo-Saxon models of capitalism in the 1990s confirmed that the socially democratic capitalisms of continental Europe were also being overtaken by history. With their market-inhibiting economic and social arrangements, the argument runs, these capitalisms are economically inferior. Many on the left drew a similar conclusion, "differ[ing] from conservative and neo-conservative critics only in deriving no pleasure

23 See M Boyco, A Shleifer and R Vishny, *Privatising Russia* (Cambridge MA: MIT Press 1997), pp. vii, 10–11. This underlies the neoliberal desire to turn as many "things" as possible into objects of private property (privatisation) capable of being transacted in markets.

24 Non-market mechanisms and institutions, like state bureaucracies, tend to be depicted as artificial, man-made, freedom-inhibiting, tendentially inefficient, market-substitutes to be deployed only in situations of "market failure".

25 H-J Chang, *Globalisation, Economic Development and the Role of the State* (London: Zed Books 2003), p. 1.

26 See, for example, Richard Posner's sudden and extraordinary conversion to the merits of state intervention: *A Failure of Capitalism* (Cambridge MA: Harvard University Press 2009); and Campbell, "The end", n. 20 above.

27 On nineteenth-century liberal attitudes to democracy, see A Arblaster, *The Rise and Decline of Western Liberalism* (Oxford: Blackwell 1984), pp. 75–9, 264–83.

28 On this, see B Jessop, *The Future of the Capitalist State* (Cambridge: Polity 2002).

29 The term the "new constitutionalism" was first coined by Stephen Gill. It is discussed later in this paper.

30 D Vittas, "The Argentine pension reform and its relevance for Eastern Europe", *World Bank Policy Research Working Paper* 1819 (Washington DC: World Bank 1997), pp. 8–9.

31 E Meiksins Wood, *The Origin of Capitalism: A longer view* (London: Verso 2002), p. 1.

from it".[32] From this perspective, the gradual evisceration of European social democracy and triumph of a neoliberal, stock- and free-market-based capitalism is more or less inevitable, a product of inexorable, transhistorical, economic laws. As Roberto Unger says, in this seeming "dictatorship of no alternatives", progressives have been reduced to trying to preserve the compensatory redistributions of the welfare state and to "humanise the inevitable".[33] After the deluge, of course, even this modest project is collapsing.

As this suggests, for neoliberals there is at work a process of functional evolution in which "the market" acts as the mechanism of selection – a process which is thought to have accelerated with the growth of increasingly open, global markets.[34] This idea is, of course, most easily applied to productive enterprises and, sure enough, in the 1990s, when the capitalisms of the US and the UK were out-performing their rivals, it came widely to be believed that this was in part because their shareholder-oriented, stock-market-centred corporations were economically superior to their more stakeholder-friendly rivals in places like Germany and Japan, leading some to suggest that we had reached "the end of history for corporate law".[35] For neoliberals, however, globalisation has also increased the competition between the legal arrangements of different societies, amplifying the "evolutionary trend towards economically more efficient legal rules". From this perspective, states are engaged in regulatory competition to produce "efficient law" and law as a whole can be "understood as the outcome of free competition among alternative, largely heterogeneous suppliers of legal authority", hence claims that law not only "determines [but] is determined by economic efficiency".[36] The concept of efficiency has thus come to be used not only to provide "an economic commentary on the law [but] . . . an explanation of [it]".[37] Here too "old Europe", hindered by its "legal origins" and insufficiently market- and investor-friendly civil law systems, is thought to fall short.[38]

Law and policy in an economically determined world

In recent decades, this economically determinist, Darwinian paradigm of development – with its curious resemblance to vulgar, base-superstructure versions of Marxism[39] – has greatly influenced policymaking. Armed with tools which, they believe, make it possible to not only plot how far different bodies of "real world law" deviate from "the efficient model" but to offer "practical directions for making laws best",[40] and firm in the belief that they are working with the grain of history, neoliberal policymakers in international agencies have displayed boundless self-confidence in dispensing policy advice to transitional

32 T Judt, "Europe vs America" (February 2005) 52(2) *London Review of Books* 37.

33 Unger, *The Left Alternative*, n. 5 above, pp. 1–11.

34 See M Roe, "Chaos and evolution in law and economics" (1996) 109 *Harvard Law Review* 641.

35 H Hansmann and R Kraakman, "The end of history for corporate law" (2001) 89 *Georgetown Law Journal* 439. In one of his (many) moments of irrational exuberance, Alan Greenspan went even further in 1998, arguing that American economic performance was so impressive, it was "possible we have moved *beyond* history": remarks before the Joint Economic Committee of Congress, 10 June 1998.

36 See U Mattei and F Pulitini, "A competitive model of legal rules" in A Breton, G Galeotti, P Salmon and R Wintrobe, *The Competitive State* (New York: Springer 1991); and U Mattei and F Calfagi, *New Palgrave Dictionary of Economics and Law* vol. I (Basingstoke: Palgrave Macmillan 1998), p. 346, at p. 351; see also Roe, "Chaos", n. 34 above, p. 642.

37 U Mattei, *Comparative Law and Economics* (Ann Arbor: University of Michigan Press 1997), p. xiii; see also R Cooter and T Ulen, *Law and Economics* 2nd edn (Reading: Addison-Wesley 1996), preface.

38 See E Glaeser and A Shleifer, "Legal origins" (2002) 107 *Quarterly Journal of Economics* 1193–229.

39 See P Ireland, "History, critical legal studies and the mysterious disappearance of capitalism" (2002) 65 *Modern Law Review* 120.

40 See Mattei and Calfagi, *New Palgrave*, n. 36 above, p. 346.

countries.[41] Neoliberal economic determinism has also greatly influenced legal scholarship, not least in fuelling a tendency to treat certain areas of law as essentially superstructural, as mere reflections or facilitators of the economic logic of "the market". This tendency has found its most vivid expression in the work of the law-and-economics movement, particularly the Posnerian variety, much of which is underlain by a belief in the transhistorical, purely economic, efficiency-maximising rationality of the "deregulated" market and in a "classical evolutionary paradigm" of legal development.[42] It is also visible in claims about "legal convergence" and the benefits of importing and imitating rules, especially in areas of law designed to facilitate trade: "legal transplants" have come to be seen as a way of accelerating evolution towards better ("more efficient") legal rules.[43] Indeed, even those who do not share the intellectual presuppositions and political preferences of neoliberalism often find themselves operating *within* rather than *against* the economically determinist analytical framework it provides. As Unger says, erstwhile progressives have become the "chastened votaries of their neoliberal opponents".[44] Many of those who favour stakeholder models of the corporation, for example, attempt to defend them on orthodox "efficiency" grounds rather than on the varied and complex grounds of individual and social well-being, justice, human dignity and fulfilment, and productive efficiency more broadly defined. Indeed, as Ronald Dore observes, there is now a tendency to view models of the corporation which do not conform to neoliberal precepts as not merely different but "abnormal", hence the emergence of a "US/UK normative–Japan/Germany deviant" mindset.[45] Within this framework, arrangements which are deemed to conform to neoliberal norms are in no need of explanation: they are assumed to be the products of natural (market) selection and the evolutionary trend towards efficiency. It is the survival of arrangements and institutions which are deemed "abnormal" and "inefficient" that needs explaining. In this context, concepts such as "path dependence" have been deployed to explain the survival of "sub-optimal" arrangements. Sustained deviations from the hallowed path of efficiency, the argument runs, are attributable to historical quirks and accidents, to entrenched cultural, political and institutional differences. Moreover, the resulting less-than-efficient arrangements can sometimes become "locked in".[46] It is also when accounting for the less-than-efficient that *power*, which is otherwise strangely absent from neoliberal explanatory frameworks, makes a belated appearance, albeit in a marginal role at the side of the stage. Thus, we have to acknowledge the possibility that "powerful interest groups" – ranging from firms, to practising lawyers, to politicians and regulators, to academics – will try to "impede the

41 This advice is, of course, associated with the so-called "Washington Consensus".

42 See Roe, "Chaos", n. 34 above.

43 See A Ogus, "Competition between national legal systems: a contribution of economic analysis to comparative law" (1999) 48 *International and Comparative Law Quarterly* 405, p. 411.

44 Unger, *The Left Alternative*, n. 5 above, p. 7.

45 R Dore, "Deviant or different? Corporate governance in Japan and Germany" (2005) 13 *Corporate Governance* 437.

46 Concepts such as path dependence – the idea that "history matters" – ostensibly soften the economic determinism and reductionism of neoliberal theories. In many ways, however, they simply reinforce them by offering an explanation for the contrary evidence that seems to be thrown up by empirical reality. In this context, so-called comparative law and economists have developed a notion of "comparative efficiency", drawing on the work of institutional economists like Oliver Williamson to argue that what is legally efficient may vary according to the institutional context in which rules operate. A particular legal arrangement may not be the most efficient one in a theoretical world, but it may well be the best that can be achieved in light of the existing formal and informal constraints. See U Mattei and A Monti, "Comparative law and economics: borrowing and resistance" (2001) 1 *Global Jurist Frontiers* 5, p. 7; and U Mattei, *Comparative Law and Economics* (Ann Arbor: University of Michigan Press 1997).

convergence of legal principles relating to homogeneous products" for personal gain. They will try to maintain barriers to competition (like restrictive choice of law rules), to "imped[e] evolution towards economically justifiable common principles" and to "resist cost-reducing reforms to the law" in "rent-seeking" attempts to secure income and wealth beyond that they would receive in a competitive market.[47]

Many elements of this analytical framework have figured in the debates surrounding pension reform. Thus pension privatisation is regularly presented as the economically rational, most "efficient" way forward, and regimes that cling to old PAYG systems warned that they are putting themselves at a competitive disadvantage. Thus, the progress made by the multi-pillar system proposed by the World Bank, Estelle James tells us, is in large part attributable to "economic forces":

> with economic growth slowing down in many countries and financial markets opening up globally, it has become increasingly important to raise productivity through improved incentives in the labor market and through the accumulation of capital which is then allocated to its most efficient uses . . .

Ideas about rent-seeking and path dependence have also been deployed, with the advocates of privatisation frequently lamenting the self-interested and economically irrational, "political" opposition to privatised systems. Thus, the extent to which individual countries have maximised the privatised pillar is said to have been significantly shaped by "political" factors such as the "strength of their social security bureaucrats, pensioner organisations and labor unions".[48]

4 The ethical case for pension privatisation: towards a world of neoliberal individuals

The case for pension privatisation rests not only on claims about its economic benefits, however, but on claims about its ethical virtues. By fostering the development of a financial property-owning "equity culture",[49] its advocates argue, pension privatisation discourages dependence and encourages individual self-reliance and responsibility, expanding the realm of freedom and choice and linking benefits to contributions and work to reward. In doing so, it helps to build ownership societies composed of autonomous, hard-working, self-sufficient, intensively possessive, property-owning, neoliberal individuals.[50] Thus, Pinera describes his ideas for social security reform as "part of an overall vision of a free market and a free society" whose goal is to enable "the whole working population [to] accumulate wealth and become shareholder-capitalists". The intention, argues Estelle James, is to "inculcat[e] the ethos of personal responsibility and [to] mak[e] each worker a little capitalist . . ."[51] From this perspective, pension privatisation is a development of potentially historic importance, part of a transformative "worker–capitalist revolution" which has to

47 Ogus, "Competition", n. 43 above, p. 411.

48 James, "New systems", n. 16 above, pp. 9, 29–30.

49 See J Poterba, "The rise of the 'equity culture': US stockownership patterns, 1989–1998" (2001), http://econ-www.mit.edu/faculty/poterba/files/aca2001.pdf.

50 Drawing on C B MacPherson, David Harvey argues that the 1990s saw a "huge shift in mental conceptions of the world such that an even more intensive possessive individualism arose . . .": Harvey, *The Enigma*, n. 3 above, p. 132.

51 J Pinera, "The bull by the horns: the battle for Chile's social security reform" (2009), www.hacer.org/pdf/Pinera04.pdf; James, "New systems", n. 16 above, p. 9. "It would be naïve", James writes, "to ignore the fact that the partial replacement of a publicly managed program by privately managed funds . . . has political and ideological implications", for "it changes the balance of power in society, taking influence and jobs away from bureaucrats and others who control the public programs and shifting them to private entrepreneurs."

overcome "philosophical" obstacles to engineer a "paradigm shift" in which collectivist ideas are abandoned and replaced by ideas of individual responsibility.[52]

From "share-owning democracies" to "ownership societies"

In the UK, the idea that pension privatisation would help to create a new world of ownership societies and neoliberal individuals first surfaced in the rhetoric of Margaret Thatcher's government. The right-to-buy scheme introduced in 1980 contributed to a rise in home ownership, and, following the privatisation of previously state-owned industries and demutualisation of building societies, there was a significant increase in direct share ownership.[53] By 1990 more than one in five households owned shares directly, compared to less than one in ten a decade earlier. Indirect financial property ownership also increased, encouraged by the running down of the state pension and the introduction of various inducements to private investment, leading the Conservative government to boast that it was "taking capitalism to the people".[54] The goal, claimed Nigel Lawson, was to encourage the creation of "an ever-widening share-owning democracy". This was reiterated by John Moore, Financial Secretary to the Treasury from 1983–86, when he argued that privatisation not only brought productive benefits but was capable of "transforming public attitudes towards economic responsibility and the concept of private property". It was "an educational process" through which people could grasp "the fundamental beliefs and values of free enterprise".[55] With its "unrivalled power to teach the responsibilities and rewards of a free society", privatisation was depicted in terms of moral regeneration.[56]

With others following the British vanguard, by the turn of the century there had been significant increases in stock market participation and financial property ownership in many parts of the world.[57] In the US, where public assets were sold off and private retirement saving encouraged, 30 million Americans became stockholders for the first time between 1989–99 and the number owning corporate stock, directly or indirectly, increased by nearly 60 per cent.[58] In similar vein, by the turn of the millennium, 17.3 per cent of households in France, Germany, Italy, the Netherlands and the UK were holding stocks directly. When indirect ownership was taken into account, the overall stock market participation rate of households rose to nearly 50 per cent in the UK, 33 per cent in the Netherlands, 23 per cent in France and 20 per cent in Germany.[59] By the time Pinera met George W Bush in 1997, he was portraying pension privatisation as a "crucial element of an ownership society". Embracing this idea, in his second inaugural speech Bush argued that to enable "citizens [to] find the dignity and security of economic independence" and to become "agents of [their] own destiny", he would "widen the ownership of homes and businesses, retirement savings

52 Skousen, *EconoPower*, n. 16 above, ch. 5; Pinera, "The bull", n. 51 above.

53 See T Clarke, "The political economy of the UK privatization programme" in T Clarke and C Pitelis (eds), *The Political Economy of Privatization* (London: Routledge 1993), p. 205.

54 J Moore, "British privatization – taking capitalism to the people" (February 1992) *Harvard Business Review* 115; See Ireland, "Shareholder primacy", n. 10 above.

55 Moore, "British privatization", n. 54 above, p. 115.

56 "Economics are the method", Thatcher argued, but "the object is to change the soul": quoted in J Gray, "Thatcher, Thatcher, Thatcher" (April 2010) 32 *London Review of Books* 19–21.

57 See L Guiso, M Haliossos and T Japelli, "Stockholding: a European comparison" in L Guiso, M Haliassos and T Jappelli (eds), *Stockholding in Europe* (Basingstoke: Palgrave Macmillan 2003), p. 201.

58 The development of defined contribution Individual Retirement Accounts (IRAs) and 401k plans were central to this development: see Investment Company Institute, *Equity and Bond Ownership in America* (Washington: ICI 2008); J M Poterba, "Stock market wealth and consumption" (2000) 14 *Journal of Economic Perspectives* 99.

59 See Guiso et al., "Stockholding", n. 57 above. British households showed the highest rate of direct ownership at 27.9%.

and health insurance".[60] It was also around this time that the ideas of "financial literacy" and "financial education" began to rise to prominence. In 2003, the OECD established an International Gateway for Financial Education, while the Financial Services Authority (FSA) launched its national strategy on financial capability. In a world in which governments are "scal[ing] back the benefits of state-supported social security programmes" and "an increasing number of workers" are having to "rely on defined contribution pension plans and their personal savings to finance their retirement", the argument runs, people need to understand finance so they can manage life's financial risks and "take ownership" of their futures. In other words, they need to start thinking like investors and recognise that achieving economic well-being in old age involves taking and managing risks.[61]

The bondholding class and the myth of democratised wealth

The number owning, or possessing a beneficial interest in, financial property has undoubtedly grown substantially in recent decades. In the US, increasing participation in defined contribution pension plans has seen the percentage of households owning stock directly or indirectly rise from about 19 per cent in the early 1980s to nearly 50 per cent.[62] The story is similar elsewhere.[63] The distribution of financial property ownership remains, however, very heavily concentrated. Indeed, in recent decades such redistributions as have occurred have tended to be upwards rather than downwards. In the US, for example, over one-third of total financial wealth is held by the wealthiest 1 per cent, with the next wealthiest 9 per cent holding another third; the bottom 50 per cent hold only 3 per cent.[64] The concentration of share ownership is even greater, with the wealthiest 10 per cent of Americans holding (directly and indirectly) over three-quarters of corporate equity and the least wealthy half of the population holding less than 2 per cent.[65] In the UK the picture is very similar: the base of the financial pyramid has widened, but financial property ownership remains very heavily concentrated among the very wealthy, with "extreme concentrations" among "the wealthiest 5 to 10 per cent of households".[66] The economist Ray Canterbery has dubbed this financial property-owning elite "the bondholding class".[67]

Pension privatisation and the spread of financial property ownership cannot, therefore, be said to have democratised wealth or generated the emergence of genuine ownership societies, even in the developed world. There is, however, clear evidence that they have prompted a shift in attitudes along the lines hoped for by John Moore. It is not insignificant

60 See Orenstein, *Privatizing Pensions*, n. 6 above, p. 1; and S Soederberg, *Corporate Power and Ownership in Contemporary Capitalism* (Routledge: Abingdon 2010), p. 27. According to Bush, "self government relies, in the end, on the governing of the self". The creation of such an "ownership society", the Cato Institute observed, required that more Americans be given "an opportunity to invest in stocks, bonds and mutual funds so that they too c[ould] become capitalists": see D Boaz, "Defining an ownership society", http:www.cato.org/special/ownership_society/boaz/html.

61 See "Round-table discussion on financial literacy", *New Statesman*, 9 January 2006.

62 It reached 49.5% in 2002, before falling back slightly to 47% in 2008: Investment Company Institute and Securities Industry Association, *Equity Ownership in America* (Washington DC: ICI/SIA 2002); Investment Company Institute, *Equity*, n. 58 above.

63 See Guiso et al., "Stockholding", n. 57 above.

64 See A Kennickell, "A rolling tide: changes in the distribution of wealth in the US, 1989–2001" (2003) *SCF Web* www.federalreserve.gov/pubs/oss/oss2/scfindex.html; J Poterba, "Stock market wealth and consumption" (2000) 14 *Journal of Economic Perspectives* 99.

65 See Ireland, "Shareholder primacy", n. 10 above.

66 See J Banks, Z Smith and M Wakefield, "The distribution of financial wealth in the UK", *Working Paper* 02/21 (London: IFS 2002); Ireland, "Shareholder primacy", n. 10 above.

67 E R Canterbery, *Wall Street Capitalism: The theory of the bondholding class* (London: World Scientific Publishing 2000). His "bondholding class" encompasses bondholders and stockholders, p. 7.

that when Lawrence Mitchell, a leading academic critic of existing American corporate governance practices, asks "who is the stockholder?", he answers "you and me".[68] The spread of financial property ownership, part of the more general "financialisation of the masses", has fostered a belief that "we're all (more or less) shareholders now", or soon will or could be. This has proved to be of considerable ideological and psychological significance, encouraging more and more people to think of themselves as middle class and as having a vested interest, albeit a modest one, in the performance of the corporate sector, financial markets and fictitious capital. This is reflected in the increasing media prominence given to financial news as indicator of the health and well-being of society as a whole.[69] Equally importantly, perhaps, with the growth in what Richard Minns has called "social security capital", there has emerged a new culture of dependency, as a growing number of workers look to, and rely on, the performance of financial markets and publicly listed corporations for security in old age.[70] These developments have played an important role in garnering support for, and legitimating, shareholder value-oriented corporations and the prioritisation of investor protection as a policy goal.

The rise of the shareholder value corporation

There is general agreement that in the post-war period, when organised labour was relatively strong and a mixture of shareholder dispersal and capital controls undermined the power and influence of financial interests, corporations became less single-mindedly shareholder-oriented. By the 1950s, even in the US and the UK, where the principle of shareholder primacy was well established, it was widely believed that corporate managers were using the discretion they had acquired over the determination of corporate goals to balance a range of different interests, rendering corporations more "socially responsible".[71] Certainly, workers *qua* workers did relatively well during this period, as reflected in falls in the levels of income and wealth inequality. From the 1970s, however, as the power of organised labour waned, financial property owners began to re-unite in institutions (including pension funds) and were liberated from the constraints imposed by Bretton Woods. Acting through their institutional representatives, they started to become more active, both within corporations and in financial markets. With the rise of the hostile takeover, emergence of an active market for corporate control and growing use of executive share options and other performance-related forms of remuneration, share price became a central focus of management. If in the 1960s and 1970s firms had, by their own accounts, been relatively insulated from investor preferences, by the 1990s they had become intensely sensitive to them. With corporate strategies increasingly tied to a narrowly financial view of how firms should be run, the maximisation of "shareholder value" emerged as the overriding corporate goal. Shareholder primacy had been reasserted with a vengeance.

The rise of the shareholder value corporation has brought significant benefits to a few. With managers under pressure to raise the market value of their corporations and to

68 L Mitchell, *Corporate Irresponsibility* (New Haven: Yale University Press 2001), pp. 147–8.

69 The bondholding class, Canterbery argues, has "revamped the indicators of economic well-being" and "redefined progress to benefit mostly themselves": *Wall Street*, n. 67 above, p. 8.

70 See R Minns, *The Cold War in Welfare* (London: Verso 2001); Soederberg, *Corporate Power*, n. 60 above, p. 36.

71 The emphasis that some still placed on "the profit-making element in corporate activity", the leading post-war British company lawyer L C B Gower argued in 1955, had "a slightly old-fashioned ring": L C B Gower, "Book review of Emerson and Latcham, *Shareholder Democracy*" (1955) 68 *Harvard Law Review* 922. On the socially responsible corporation, see C Kaysen, "The social significance of the modern corporation" (1957) 47 *American Economic Review* 311; and T Nichols, *Ownership, Control and Ideology: An inquiry into certain aspects of modern business ideology* (London: George Allen & Unwin 1969). The resulting dilution of the principle of shareholder primacy was widely thought defensible as corporations resembled social or quasi-social institutions rather than purely private enterprises, while shareholders looked less and less like proper owners.

distribute a larger proportion of profits as dividends, the share of national income accruing to financial property owners rose across the OECD.[72] In the US, as the policy of "retain and invest" was replaced by one of "downsize and distribute", the proportion of post-tax income distributed to shareholders significantly increased.[73] In the UK, there was an even more marked upward shift in pay-outs from 13 to 20 per cent in the 1980s to 20 to 35 per cent in the 1990s and early 2000s.[74] Further shareholder gains followed the rises in the stock-market value of shares, though many of these proved transient. These benefits not only came at the expense of labour, which suffered as corporations downsized, outsourced and "re-engineered", but were concentrated among the bondholding class and a coterie of corporate executives and financial intermediaries. The losses suffered by workers *qua* workers significantly outweighed the gains they made as worker–capitalists.

Shareholder value corporations have nevertheless been vigorously defended, though not so much on the problematic basis of shareholder corporate ownership as on the basis of their superior efficiency.[75] In this regard, the arguments made for shareholder value corporations echo those for made for pension privatisation. Claims that they operate for the benefit of society as a whole have, of course, been given further credence by the growth in private pensions and emergence of workers as small-scale investors. Thus, in seeking to account for the growing support for shareholder-oriented corporations, Hansmann and Kraakman, in their neoliberal tour-de-force announcing the "end of corporate history", point not only to the "important economic forces" operating in their favour but to the rapid expansion of equity ownership and emergence of a "public shareholder class". Echoing Pinera, they argue that with "even blue-collar workers now often hav[ing] sufficient personal savings to justify investment in equity securities . . . labor and capital no longer constitute clearly distinct interest groups in society".[76]

The prioritisation of investor protection

The claim that shareholder value corporations and open financial markets operate to maximise efficiency, wealth and welfare, coupled with the feeling that "we're all (more or less) shareholders (or potential shareholders) now", has also helped to legitimate the growing prioritisation by policymakers of property protection and particularly the protection of vulnerable intangible property forms like intellectual and financial property. In recent decades, ensuring the integrity of financial property – investor protection – has become one of the principal policy goals of governments, routinely trumping other economic and social objectives. Nowhere has this been clearer than in the policy responses to both the initial financial crisis and the subsequent crises in sovereign debt, where the overriding concern has been to try to ensure that financial property owners continue to

72 G Epstein, "Introduction", and G Epstein and A Jayadev, "The rise of rentier incomes in OECD countries", both in G Epstein (ed.), *Financialization and the World Economy* (Cheltenham: Edward Elgar 2005).

73 W Lazonick and M O'Sullivan, "Maximising shareholder value: a new ideology for corporate governance" (2000) 29 *Economy and Society* 13.

74 J Froud, S Johal, A Leaver and K Williams, *Financialization and Strategy: Narrative and numbers* (London: Routledge 2006), pp. 68, 87–8.

75 Legally speaking, it is very difficult to sustain the "ownership" claims of the shareholders of large publicly quoted corporations: see P Ireland. "Defending the rentier: corporate theory and the reprivatization of the public company", in J Parkinson et al. (eds), *The Political Economy of the Company* (Oxford: Hart 2001), p. 141.

76 Hansmann and Kraakman, "The end of history", n. 35 above. This class, they argue, operates as a coherent, "broad and powerful interest group in both corporate and political affairs across jurisdictions". The new Chilean pension system was inaugurated on Labour Day, 1 May 1980, to highlight that in a free-market economy the interests of capital and labour are "convergent": see Skousen, *EconoPower*, n. 16 above, ch. 5.

receive their revenues whatever the cost. As we have seen equally clearly, however, protecting financial property is not easy.

This is because the value of intangible financial property is derived not from its physical properties but from its anticipated future earning power.[77] As such, its value is inherently speculative,[78] rendering it unusually vulnerable not only to fraud, bubbles and the manipulation of expectations, but to changes in trade and labour regimes, in financial and taxation policies, as well as in general economic conditions and the balance of political and class power. It is a high maintenance property form, the preservation of whose integrity and value demands an extraordinarily wide range of legal and other interventions aimed at ensuring that the revenues continue to flow.[79] Indeed, it was the desire to hedge and spread the risks associated with financial property that prompted some of the financial innovations – securitisation, credit default swaps and the like – which Alan Greenspan thought had made "individual financial institutions . . . less vulnerable to shocks [and] the financial system as a whole . . . more resilient", but which we now know exacerbated the scale of the collapse.[80]

The growing importance attached to investor and financial property protection is evident in the macro-economic focus on low inflation,[81] in the growing willingness of both states and international organisations to act as financial firefighters, as well as in the rise of regulatory bodies, such as the FSA, and rapid growth in laws regulating securities and capital markets.[82] It is also evident in the emergence of a complex transnational framework aimed at securing the interests of investors and of foreign investors in particular. In recent years, for example, international institutions have prioritised the creation of "good climates for investment", identifying 12 areas where "standards are important for the institutional underpinning of macroeconomic and financial stability" and developing universal standards of "good governance" in each.[83] The OECD's Principles of Corporate Governance, for instance, seek to embed the Anglo-American shareholder-oriented model of the corporation around the world and to protect the interests of *rentier* shareholders. Although compliance with these standards is formally voluntary, in reality it is more or less compulsory if capital flight and investment strikes are to be avoided.[84]

At the same time a new transnational "investment rules regime" has been constructed. Embodied in a variety of legally binding agreements, ranging from bilateral investment treaties to regional trade agreements to multilateral instruments, this regime provides various legal protections to foreign investors. In addition to insisting on non-discrimination and fair and equitable treatment, it inhibits certain sorts of state interventions, prohibiting

77 More specifically, from a capitalisation of the returns which are expected to accrue to it in the future.

78 This is most obviously true of corporate shares, the rate of return on which is not usually fixed in advance but varies according to the profitability of the company. More recently, however, the uncertainties have encompassed the sovereign debt of a growing number of countries.

79 See P Ireland, "Property and contract in contemporary corporate theory" (2003) 23 *Legal Studies* 453. As Gillian Tett observes, it is not for nothing that the root of the word credit comes from the Latin *credere*, meaning to believe: see "Lost through creative destruction", *Financial Times*, 10 March 2009.

80 A Greenspan, speech to American Bankers Association Annual Convention, New York, 5 October 2004.

81 See Canterbery, *Wall Street*, n. 67 above: "Because of the bondholders' primal fear of inflation, monetary policy has had a strong anti-inflationary, anti-growth, and anti-employment bias", p. 7.

82 Exemplified by things such as the EU's Markets and Financial Instruments Directive of 2004: see N Moloney, *How to Protect Investors: Lessons from the EC and the UK* (Cambridge: CUP 2010).

83 See the Financial Stability Board, *Compendium of Standards*, available at www.financialstabilityboard.org/index.html.

84 On this see S Soederberg, *The Politics of the New Financial Architecture* (London: Zed Books 2004).

not only measures that "directly or indirectly" expropriate investment interests (such as nationalisation) but so-called "creeping expropriations" or "regulatory takings" – state interventions, like corporate or environmental regulations, that adversely impact on the value of investments. With investors increasingly being given standing to sue if their investments are damaged, states are becoming more constrained not only by fears of capital flight but by the threat of litigation.[85] This "interlocking network of rules and rule-making structures" thus offers investors quasi-constitutional protections, insulating them from "the vicissitudes of democratic politics".[86] This "new constitutionalism"[87] locks countries into a new "rule of (investment) law" which not only protects foreign investors from legislative and administrative actions that damage their investment interests but confers on international organisations powers which, when exercised by states themselves, are usually referred to as "sovereign powers".[88]

In this context, as in others, the gulf between neoliberal rhetoric and neoliberal practice is palpable. As many have observed, you simply cannot understand the changes in the global financial markets since the early 1970s in terms of deregulation. On the contrary, financial markets have become increasingly heavily regulated. Indeed, modern American financial markets are probably "the most highly regulated markets in history if regulation is measured by volume (number of pages) of rules . . . and by extent of surveillance", and "possibly even by vigour of enforcement".[89] In pursuit of investor protection, states have endorsed extensive (regulatory) interventions in some areas, while at the same time reducing, discouraging and prohibiting them in others.

5 The neoliberal vision unravels

Pension privatisation, corporate governance and financialisation

Has pension privatisation delivered on its promises? Despite the claims made in its favour, there is little evidence that pension privatisation has fostered economic growth. This is not, perhaps, surprising, for, as John Eatwell observes, "exhaustive studies have identified no impact of increased personal savings on the rate of investment".[90] The growth in pension saving has, however, undoubtedly fuelled both increased investment in the claims on productive activity (in financial property) and the growth in financial transactions and financial markets. In doing this, it has contributed less to economic growth than to the processes whereby, to use Keynes' phrase, "speculation has come to dominate

85 In both domestic courts and international tribunals: See D Schneiderman, *Constitutionalizing Economic Globalization: Investment rules and democracy's promise* (Cambridge: CUP 2008); D Schneiderman, "Investment rules and the new constitutionalism" (2000) 25 *Law and Social Inquiry* 757; D Schneiderman, "Constitutional approaches to privatization: an inquiry into the magnitude of neo-liberal constitutionalism" (2001) 63 *Law and Contemporary Problems* 83; D Schneiderman, "Investment rules and the rule of law" (2001) 8 *Constellations* 521.

86 See Schneiderman, "Investment rules and the rule of law", n. 85 above, p. 522; Schneiderman, *Constitutionalizing Economic Globalization*, n. 85 above, pp. 3–4; see also R Hirschl, "The political origins of the new constitutionalism" (2004) 11 *Indiana Journal of Global Legal Studies* 71–3. See also D Sarooshi, *International Organizations and their Exercise of Sovereign Powers* (Oxford: OUP 2005).

87 A term coined by Stephen Gill: see his "Globalisation, market civilisation, and disciplinary neoliberalism" (1995) 24 *Millennium: Journal of International Studies* 399–423; and "The constitution of global capitalism" (paper presented at the ISA Annual Convention, Los Angeles, 2000).

88 See Schneiderman, "Investment rules", n. 85 above (2001); and also Sarooshi, *International Organizations*, n. 86 above.

89 D MacKenzie, "Opening the black boxes of global finance" (2005) 12 *Review of International Political Economy* 569. See also Campbell, "The end", n. 20 above, p. 324.

90 This leads him to conclude that "all those arguments about the need for society to save more are quite simply poppycock": J Eatwell, "Three fallacies on pensions" (2003), www.cerf.cam.ac.uk/publications/files/Eatwell-%20pensions%20crisis.pdf.

enterprise".[91] There is equally little evidence that the growing power of pension funds and other institutional investors has generated significant productive benefits at the level of the individual corporation. On the contrary, while it has helped to ratchet up the financial and capital market pressures on managers, this has merely served to reinforce their obsession with share price and other financial indicators of performance. With more and more securities analysts providing institutions (including pension funds) with information on which to base their decisions, corporations have become increasingly concerned with keeping the expectations of analysts in line with their own forecasts, constructing plausible narratives about their strategy and performance and delivering numbers which corroborate their stories, even if this entails using accounting and revenue manipulations of various sorts.[92]

In similar vein, while the takeovers associated with the market for corporate control have brought quick and plentiful gains for their organisers and target company shareholders, the overall verdict of studies of their long-term impact on corporate performance is generally negative. A significant proportion of the profits generated by hostile takeovers, including those associated with the private equity inspired boom of recent years, have been derived from asset stripping aimed at boosting short-term yields and at creating the illusion of improved productivity, rather than genuine improvements in output and productivity.[93] As Froud et al. observe, "for all its prestige" the business model of private equity firms "is almost exactly like used-car trading", where capital is borrowed, cars purchased and cosmetically fixed up, before being resold for a quick profit.[94] Endorsing this view, Robert Peston emphasises the need to distinguish private equity firms from "real venture capitalists" and "genuine entrepreneurs". The former, he argues, tend to view firms "in a very impersonal and blinkered fashion", as "property and chattels, and statistics about cash flows and market shares" rather than as productive enterprises. They are, he concludes, better at financial engineering than successfully running businesses in the long-term, showing "little empathetic understanding of a business as a social institution wholly dependent on its people". Indeed, for workers, their activities are often disastrous, as they ruthlessly sweat all assets, human as well as capital.[95]

As this suggests, the effects of the growing capital market pressures on corporate culture have been considerable. Many non-financial corporations have become quasi-financial corporations and more and more corporate managers have begun to behave like financial market participants who see their enterprises as little more than financial flows and bundles of assets capable of being bought and sold at will. Production is often increasingly incidental to the much more lucrative business of "balance-sheet restructuring".[96] It is hardly surprising that many think the growing focus on "shareholder value" has had a corrosive effect on productive performance and on innovation. Nor that many believe the

91 J M Keynes, *The General Theory of Employment, Interest and Money* (London: Macmillan 1936): "Speculators may do no harm as bubbles on a steady stream of enterprise. But the position is serious when enterprise becomes the bubble on a whirlpool of speculation." More generally, see J B Foster, "The financialization of accumulation" (October 2010) 62(5) *Monthly Review*.

92 See Froud et al., *Financialization*, n. 74 above.

93 S Konzelmann, F Wilkinson and M Fovargue-Davies, "Governance, regulation and financial market instability: the implications for policy" (paper delivered to Re-engineering the Corporation Seminar, Birkbeck College, 26 March 10), pp. 15, 19.

94 Froud et al., *Financialization*, n. 74 above, p. 122.

95 R Peston, *Who Runs Britain?* (London: Hodder 2008), pp. 13–14, 22, 44–5, 59–61, 96. Paradoxically, corporate raiders of this sort have often found willing allies in pension fund managers seeking to maximise returns for their (worker) investors.

96 J Toporowski, "The wisdom of property and the politics of the middle classes", (2010) 62(4) *Monthly Review*.

growing bias towards quick short-term gain rather than new, long-term productive investment has seen resources poorly allocated from a productive and social perspective. In recent years, more and more businesspeople, journalists and academic commentators have expressed concern about the largely negative effects of an increasingly financialised governance culture.[97]

Pension privatisation and the financial crisis

Pension privatisation has also failed to deliver a world of "ownership societies". Even before the recent financial crisis, pensioner poverty around the world was rising, especially amongst women, and private pension schemes were facing serious financial difficulties. In the UK between 1992 and 2004, the coverage of workers by final salary, defined benefit schemes – theoretically less vulnerable to stock market volatility than their defined contribution counterparts – fell from 21 per cent to 9 per cent. Since then, many such schemes have reported massive funding shortfalls; by 2008 only one in five were still open to new members.[98] As a result, British workers are increasingly relying on defined contribution schemes, many of which are themselves in difficulty. In March 2009, an analysis of market returns by PwC suggested that people who had been paying into such schemes for 20 years were no better off than they would have been had they left their money in cash savings accounts.[99] It is a similar story elsewhere. In Chile, for example, dissatisfaction with the privatised system emerged as a key election issue in 2006 and in Argentina in 2008, in a measure which commanded widespread popular support, the government nationalised its private pension funds, prompting the OECD to fret that others might follow Argentina's lead.[100]

The financial crisis has greatly exacerbated these problems, "wreak[ing] havoc on retirement plans of all varieties", particularly employer-based and private retirement savings schemes.[101] The market value of and returns accruing to many forms of financial property have fallen, slashing the value of pension pots and generating (or amplifying) pension fund shortfalls. Many who thought they had provided for comfortable retirements are having to think again. Many of the companies sponsoring defined benefit plans have become insolvent; many others have closed their schemes and are now cutting their contributions to the defined contribution schemes which replaced them. In recent months. UK public sector pensions, many of which offer only very modest retirement incomes, have come under

97 See, for example, J Crotty, "The neoliberal paradox: the impact of destructive product market competition and 'modern' financial markets on nonfinancial corporation performance in the neoliberal era", in Epstein, *Financialization*, n. 72, p. 77. By 2007, even Martin Wolf, erstwhile champion of free markets and globalisation, was worrying about the effects of "unfettered finance" and arguing that the new "global financial capitalism" had brought "the triumph of the global over the local, of the speculator over the manager and of the financier over the producer": M Wolf, "The new capitalism", *Financial Times*, 19 June 2007; "Risks and rewards of today's unshackled global finance", *Financial Times*, 27 June 2007.

98 See R Minns, "Missing the point" (paper delivered to Financial Institutions and Economic Security Conference, May 2009).

99 Worse still, people who had been paying in for only 10 years would now be holding less in their retirement accounts than the total value of their contributions, see *Financial Times*, 3 March 2009.

100 Since their creation in 1994, the privately owned funds had saved 96,000 million pesos for retirement accounts while pocketing 36,000 million pesos in fees. See P Antolin and F Stewart (2009), "Private pensions and policy responses to the financial and economic crisis", *OECD Working Paper on Insurance and Private Pensions* No 36 (France: OECD), p. 5, regretfully reporting that "there are policy discussions about reverting back towards PAYG public pensions in some central and Eastern European countries".

101 C O'Murchu et al., "The pensions crisis", 27 May 209, FT.com, video and audio interactive graphics. Matters have been made worse by the fact that in many countries government debt has mushroomed and both future pension savings and future government revenues are being adversely affected by falling profits and rising unemployment.

fierce attack from both politicians and the press for being too generous ("gold-plated"), prompting the government to inflation-link them not to the retail prices index but to the historically lower consumer prices index; a similar change to the inflation indexing of private pensions has been announced. Even die-hard supporters of pension privatisation, usually noted for their sublime optimism, have conceded that "the financial turmoil and the ensuing financial crisis have had a major impact on private pension assets". The OECD estimates that there were declines of between 20 and 25 per cent (over $5 trillion) in global pension assets during the course of 2008 alone.[102]

Fetishised money: the pensions illusion

In short, the financial crisis has highlighted the fundamental flaws in a pensions policy which relies on mass ownership of *rentier* financial property, making it painfully clear that private pensions are incapable of providing security in old age for more than a privileged minority. It has exposed the neoliberal vision of a world of classless ownership societies populated by autonomous, self-reliant, financial property-owning worker–capitalists as a chimera based on a series of misconceptions about the nature of financial property and money.

These misconceptions loom large in the work of those advocating pension privatisation. The World Bank, for example, regularly equates increased pension saving with increased productive investment, and stock market growth with economic growth. "Real" investment in productive enterprise is muddled up with investment in the claims on the product of such enterprise. Growth in the volume and value of interest- and dividend-bearing financial property and in the size of financial markets is treated as synonymous with growth in real productive assets and capacity. This reflects the deeply rooted tendency, endemic to modern capitalism, to treat money and financial property as somehow autonomously productive, as if financial property were capable of making more money by itself. This tendency surfaces when advocates of "ownership societies" suggest that pension privatisation will prevent the old from becoming a burden on the state and future generations. If everyone saved more and there was a move from PAYG towards fully funded pensions, they suggest, the younger generation would be relieved of the burden of supporting the old, for the latter would be able to live off the revenues accruing to their financial property. Even those critical of pension privatisation tend to make this assumption. Thus, Orenstein argues that, while privatised pension systems face certain risks social security systems do not, under privatised systems the demographic risks to which social security systems are vulnerable are "substantially reduced, because pension benefits do not rely on the earnings of another generation".[103]

There have been few better expressions of the fetishised view of financial property underlying these ideas than an advertisement which appeared in the UK in the 1980s. Depicting a man lying on a sofa, it declared that, contrary to appearances, he was working because he had been wise enough to invest in a particular financial institution. As a result his money was not only working for him but doing so incessantly because it "never slept"; it was always earning interest and dividends. On the face of it, the man was self-reliant and a burden on no one. In reality, of course, money is not autonomously productive and even if the sofa-bound man was *not* working, someone, somewhere was and it was these unknown others who were generating the wealth to which the man was entitled to lay (partial) claim by virtue of his financial property ownership.

102 See Antolin and Stewart, "Private pensions", n. 100 above. The falls have been especially large in countries, such as the US, where portfolios contain a higher proportion of equities.

103 Orenstein, *Privatizing Pensions*, n. 6 above, p. 184.

The reality is that no matter how much financial property (money) they own, the old are always ultimately drawing, one way or another, on the labour of those who work. As John Eatwell observes, people's standards of living are sustained by a flow of goods and services, which during their working lives is usually sustained by paid work. After retirement the flow can be sustained by either squirreling away goods and services and then gradually consuming them – a method with serious drawbacks, not least because many services, such as health care, simply can't be stored – or by acquiring monetary claims – revenue rights derived from either financial property and/or public pensions – which can be used as and when needed to purchase the goods and services being produced by the current workforce. Either way, the provision of goods and services for the old always involves an intergenerational transfer.[104] No matter how pensions are financed, the old are always drawing on the productive activities of the young, always making claims on the product of the labour of others. Even private pensions involve intergenerational transfers.

The unavoidable dependence of the non-working old on goods and services produced by current workers prompts Eatwell to argue that "in overall macro-economic terms" there is no difference between PAYG and so-called fully funded, privatised pension schemes so far as overall intergenerational transfers are concerned. "The burden on the current workforce, defined as the goods and services that are 'extracted' from them, is exactly the same" under both systems. Privatised schemes simply extract the resources in a different way, though in doing so they alter the distribution of goods and services between pensioners, favouring those who have invested in financial property. Comparisons between the two schemes must, therefore, be made on other grounds and in this regard, Eatwell suggests, fully funded, privatised schemes suffer from "rather weighty disadvantages", most notably their regressive impact on the distribution of pensioner income[105] and high administration costs. Public PAYG pension schemes are simple, transparent and have administration costs of only 3 to 4 per cent, compared to 20 per cent or more for the typical private scheme.[106] This leads Antonio Tricarico to conclude that the "single-mindedness of the World Bank in promoting privatised systems has been peculiar, since the evidence – including that documented in World Bank publications – has indicated that well-run public sector systems, like the social security system in the US, are far more efficient than privatised systems".[107]

Eatwell is similarly sceptical of claims that by encouraging investment and capital market development privatised pensions foster good corporate governance and the efficient allocation of resources. "Whilst it may be possible to argue that the existence of fully funded schemes promotes the development of financial markets", he argues, "there is no clear relationship between the growth of financial markets and aggregate savings growth or economic efficiency".[108] On the contrary, as we have seen, the growing power of finance has fostered the development of highly financialised and dysfunctional forms of corporate governance. As many commentators have pointed out, world growth rates have fallen during the neoliberal era, from an average of 4.8 per cent between 1960 and 1980 to

104 J Eatwell, "The anatomy of the pensions crisis" in UN Economic Commission for Europe, *Economic Survey of Europe* No 3 (Geneva: UNECE 1999). As Morris Cohen observed: "The owners of revenue-producing property are in fact granted by the law certain powers to tax the future social product.": "Property and sovereignty" (1927-28) 13 *Cornell Law Quarterly* 8, p. 13.

105 Between classes and between men and women.

106 Eatwell, "The anatomy", n. 104 above.

107 See A Tricarico, in *Social Watch Report 2007* (Montevideo: Instituto del Tercer Mundo 2007). The fees and commissions associated with privatised systems come directly out of the money that retirees would otherwise receive, "lowering their retirement benefits by as much as one third compared to a well-run public social security system".

108 Eatwell, "Three fallacies", n. 90 above, and "The anatomy", n. 104 above.

2.9 per cent between 1980 and 2000; growth in labour productivity fell from 2.5 per cent to 0.8 per cent during the same period.[109] "Perhaps the greatest advantage of fully-funded schemes", Eatwell concludes, is that they generate "an automatic adjustment of the level of pensions to the available resources" without any need for an "overt political decision". Pensioners may be disappointed with their pensions, but "they do not perceive any deliberate political decision in the reduction of their pensions by inflation or by the failure to attain a suitable return in the financial markets". In other words, private pensions depoliticise the level of provision.[110]

The class dimensions of financial property and investor protection

As Eatwell's comments suggest, the advocates of pension privatisation tend to overlook not only the intergenerational dimensions of all forms of pension provision, but the *class* dimensions of all forms of financial property. The latter were implicitly recognised by Aristotle, Aquinas and other early critics of usury when they argued that money is inherently sterile and unproductive and that all interest payments therefore inevitably entail transfers, in money form, of part of the product of one person's labour to another. The usurer, like the man on the sofa, makes money by taking, in the form of interest, part of the product of someone else's labour. When the rise of capitalism generated the development of an increasingly sophisticated and complex credit system and a huge expansion in the forms and volume of interest-bearing revenue rights, Marx saw in the modern owner of financial property, the money capitalist, a contemporary and class-based version of the usurer. For Marx, the only significant difference between the usurer's capital and the interest-bearing capital of the money capitalist lay not in their form (both entail a movement from M-M1) but in the social and class relations within which the money moved. Crucially, he argued, although the revenues accruing to money capital are all ultimately derived from the exploitation of labour and the appropriation of surplus value in the process of production, under capitalism money appears to be autonomously productive, "a mysterious and self-creating source of interest". It was because the revenues accruing to money capital seemingly presupposed no production at all (M-M1) that Marx considered it to be the most fetishistic form of capital of all.[111]

Marx did not, however, anticipate the growing integration of the working classes into financial relations. Recent decades have seen the rise of working-class ownership of financial property as well as a massive growth in working-class consumer and household (including mortgage) debt and the consequent re-emergence of new, highly developed forms of usurer's capital.[112] With the rise of "equal opportunities financialisation", mainstream financial institutions have "moved into areas they had previously redlined" and adopted a similar position "vis-à-vis the poor [as] the loan sharks and illegal lenders whose role reformers had sought to undercut".[113] In an increasingly insecure world of stagnating real wages and diminishing public provision, the desire of people to participate in the ownership society and to find security through asset ownership has been ruthlessly

109 World Bank, *World Development Indicators* (Washington DC: World Bank 2005).

110 Eatwell, "Three fallacies", n. 90 above.

111 Karl Marx, *Capital*, vol. 3, chs 25 and 29. See also Rudolf Hilferding, *Finance Capital* (1909)

112 "It was the usurious and exploitative character of lending to wage-earning households, as it developed in the concrete setting of rising inequality and privatization of the past three decades, that made it highly profitable.": see P dos Santos, "At the heart of the matter: household debt in contemporary banking and the international crisis", *Research on Money and Finance Discussion Paper* 11 (London: SOAS May 2009), p. 28.

113 L Panitch, M Konings, S Gindin and S Aquanno, "The political economy of the subprime crisis" in L Panitch and M Konings (eds), *American Empire and the Political Economy of Global Finance* (Basingstoke: Palgrave Macmillan 2008), p. 253, at pp. 264, 268, 285.

encouraged and exploited. Indeed, the whole sub-prime superstructure rested on predatory lending and regular payments of interest by the poor. Through processes of this sort, much profit-making and surplus labour extraction in the developed world has come to centre on debt rather than exploitation in production itself. Thus, one of the most significant aspects of recent changes in banking activities has been the shift in focus from productive enterprise towards the individual wage income of ordinary people as sources of profit.[114] The result of these processes has been the emergence of a growing number of worker–capitalists, of people who straddle the class divide in that they are simultaneously both workers, part of the product of whose labour is appropriated either in production or through debt, and the owners of (small amounts of) financial property which entitles them to receive interest payments on retirement. There is thus more than a little truth in the claims that what we have seen in recent years is not only the financialisation of society but the financialisation of class power.[115]

These developments have blurred class divisions but done little to eradicate them. As we have seen, while privatisation and pension privatisation in particular have undoubtedly widened financial property ownership, they have done little to change its distribution. Indeed, the alleged democratisation of financial property ownership and widening of the "community of money" has been accompanied by growing income and wealth *in*equalities. The significance of this was highlighted by the German economist Margrit Kennedy when she calculated the interest paid and interest received by the 10 income deciles of the German population. The bottom eight deciles all paid more in interest than they received; the ninth received slightly more than it paid and the tenth (and richest) received about twice as much as it paid. Much of the interest gain was concentrated in the top 1 per cent. This, she suggested, "explains graphically, in a very simple and straightforward way, why the rich get richer and the poor get poorer".[116]

As a result of the sharply skewed distribution of financial property ownership, the rise of the shareholder value corporation and growing policy focus on investor protection have brought substantial benefits to the bondholding class but few benefits to anyone else. As we have seen, even for those privileged members of the working class with private pensions, financial property ownership does not guarantee a financially secure retirement. For the much more numerous, less privileged, non-financial-property-owning members of the global working class the benefits have been even harder to discern. Indeed, as we have seen, the impact of the growing financial power on the lives of workers *qua* workers has been extremely damaging. Wage bills have been cut, real wages have stagnated, job insecurity has increased, long-hours cultures have emerged, and working practices and labour markets have been made more "flexible".[117] As Doug Henwood says, workers have been asked "to trade a few extra percentage points return on their pension fund, on which they may draw some decades in the future, for 30 or 40 years of falling wages and rising employment insecurity".[118]

---

114 dos Santos, "At the heart", n. 112 above, pp. 5–7. Costas Lapavitsas calls the extraction of financial profit directly out of the personal income of workers "financial expropriation": see his "Financialised capitalism: crisis and financial expropriation" (2009) 17 *Historical Materialism* 114.

115 Panitch et al., "Political economy", n. 113 above, at p. 257.

116 M Kennedy, *Interest and Inflation Free Money* (Philadelphia: New Society Publishers 2005).

117 In the words of one commentator, the "serial restructuring" which has accompanied the worship of shareholder value has "elevate[d] breach of implicit stakeholder contract into a guiding principle of management": Froud et al, *Financialization*, n. 74 above, pp. 109–36.

118 D Henwood, *Wall Street* (London: Verso 1997), p. 293.

As noted earlier, however, other elite groups have also benefited from financialisation. Pension privatisation, for example, has been very good for financial institutions, generating all manner of fees and commissions. At the time of the 2006 Chilean election, the privately run pension funds had over a five-year period recorded average annual profitability of more than 50 per cent. A World Bank study concluded that they were retaining between a quarter and a third of workers' contributions in commissions and fees.[119] In similar vein, lending in the US sub-prime debacle was driven less by the realistic ambitions of poor people to become homeowners and more by predatory money-making by banks. Mortgage lenders began to do everything they could to sign up borrowers at above average, sub-prime interest rates, intending to pool, securitise and sell the debt on as tranches of various grades of collaterised debt obligation.[120] In the corporate context, the most visible beneficiaries of financialisation have been corporate executives, whose remuneration has skyrocketed, often without any meaningful connection to performance. There have been unprecedented increases in executive–worker compensation ratios, especially in places such as the US and the UK.[121] Many other less visible groups have also benefited from these developments – a diverse bunch of corporate advisors and service providers, securities analysts, hedge-fund operators, private equity firms, city lawyers, and investment banks. Some of them are largely reactive, responding to corporate demands; others are proactive deal-makers, actively initiating mergers, acquisitions and financial innovations.[122] These groups have a stake in an "economy of permanent restructuring", making money from advising, trading, dealing and investing, from acquisitions and de-mergers, new issues, buybacks, securitisation and the re-bundling of risks. Indeed, they have been largely responsible for the hyper-innovation that has produced billion-dollar turnovers in capital market dealings.[123] They also, some argue, helped to develop the idea of "shareholder value" to justify their self-serving activities.[124] In short, financialisation has generated the emergence of "new layers of *rentiers*" who profit from financial expropriation, drawing substantial incomes from their positions within the financial system rather than from direct ownership of financial property.[125]

Crucially, these various elite groups, whose interests are by no means entirely co-extensive or harmonious, now wield, as various commentators have observed, considerable political power and have come in many places to dominate the state itself. In the UK, for example, the City has been given more or less everything it wants by successive governments. "Guiltlessly rapacious and mentally pugnacious", writes Hywel Williams:

> Britain's financial and business elites at least display the virtue of candour about their ultimate goals: the making of money for themselves . . . The City has won all the necessary battles for command and control. It now absorbs and directs the aims of all other power elites and thereby makes those elites subordinate to its own interests.[126]

119 "Chile's candidates agree to agree on pension woes", *New York Times*, 10 January 2006.

120 J Lanchester, *Whoops* (London: Allen Lane 2010), p. 103.

121 The 1990s, with its bull-market, is often portrayed as a decade in which managers armed with stock options heroically created value for shareholders, but in reality they were for the most part simply enriching themselves: Froud et al, *Financialization*, n. 74 above, pp. 54–64, 94

122 I Erturk, S Johal, A Leaver and K Williams, *Financialization at Work* (London: Routledge 2008), pp. 26–9.

123 P Folkman, J Froud, S Johal and K Williams, "Working for themselves?: capital market intermediaries and present day capitalism" (2007) 49 *Business History* 552; see also F Dobbin and D Zorn, "Corporate malfeasance and the myth of shareholder value" (2005) 17 *Political Power and Social Theory* 179.

124 D Zorn, F Dobbin, J Dierkes and M S Kwok, "The new new firm: power and sense-making in the construction of shareholder value" (2006) *Nordiske Organisationsstudier* 3.

125 See Lapavitsas, "Financialised capitalism", n. 114 above, p. 114.

126 H Williams, *Britain's Power Elites* (London: Constable 2006), p. 215; See also Peston, *Who Runs Britain?*, n. 95 above.

In innumerable ways and innumerable areas, from education to health to welfare, "bottom-line City imperatives have been transplanted wholesale into British society".[127] Indeed, it is difficult to imagine a more powerful testament to the growing power (and prioritisation of the interests) of the bondholding class and these financial elites than the recent bail-outs and enforced austerity. States have, in effect, decided that. as far as possible, ordinary people, rather than the bondholding class, should suffer the consequences.[128]

6 Neoliberalism in the age of austerity

Neoliberalism as a class project

Neoliberalism presents itself as committed to "rolling back the state", as wanting to allow free rein to the market. This has led many commentators to depict the current crisis as a crisis of deregulation and a very particular Anglo-American model of capitalism and to seek a solution in more and better regulation, in regulatory reform and a redrawing of the boundaries between state and market. While there is some truth in this view, it takes neoliberalism's ideological self-representation too seriously. Neoliberalism has indeed entailed some deregulation, not least in the financial sphere, but, in general, what we have seen in recent decades is not so much a retreat as a major change in the nature and forms of state activity.[129] In many spheres, there has been a marked increase in interventionism, not only by states but by international agencies, much of it, as we have seen, directed at creating "good climates for investment". The financial crisis occurred in a heavily regulated world and those who seek to understand the neoliberal era only at the ideological level, in terms of its self-professed determination to free markets from states, will struggle to make sense of actual neoliberal practice.

To understand the gulf between neoliberal practice and rhetoric, it is to the class dimensions of neoliberalism and financial property that one needs to turn. This is not to say that all the policy changes associated with the neoliberal revolution can directly or adequately be grasped through a simple class prism. As the intellectual grip of neoliberal ideas has tightened, they have taken on a life of their own. However, as a number of commentators have argued, many aspects of the neoliberal revolution are best understood as parts of a class project which emerged out of "the long downturn", the decline in growth rates and rate of return on capital investment which began in the 1960s and saw a significant fall in the share of the social product accruing to financial property owners.[130] From this perspective, its deregulatory dimension, such as it is, has entailed not so much a withdrawal of the state from economic affairs as a state-led restructuring of economic life. Many neoliberal policies have sought to reshape, in favour of capital, what Robert Hale called the "structures of mutual coercion" which characterise all markets. Law has been central to this project, for, as Hale observed, the legal rights possessed by market actors are key

127 D Kynaston, *City of London: A club no more* vol. 4 (London: Chatto & Windas 2002), p. 791.

128 See E Moya, "How the bond vigilantes punished the sick men of Europe", *The Guardian,* 18 November 2010. As Moya points out, the bond vigilantes who are exerting the market pressures which are prompting governments to slash spending and jobs are (paradoxically) primarily pension funds. She concludes that "the vigilantes are now arguably more powerful than governments", inevitably bringing to mind Bill Clinton's rhetorical question: "You mean to tell me that the success of the economic program and my re-election hinges on the Federal Reserve and a bunch of fucking bond traders?"

129 B Jessop, *The Future of the Capitalist State* (Cambridge: Polity 2002).

130 See Harvey, *A Brief History*, n. 17 above; R Brenner, *The Economics of Global Turbulence* (London: Verso 2005); R Brenner, "Interview" (March–April 2009) *Against the Current*, issue 139; R Brenner, "What is good for Goldman Sachs is good for America" (Los Angeles: UCLA, Institute for Social Science Research, April 2009); G Dumenil and D Levy, *Capital Resurgent* (Cambridge MA: Harvard University Press 2004).

determinants of the coercive power they can exercise.[131] Put simply, in recent decades the coercive power of capital has been greatly enhanced by (amongst other things) multiple extensions to its legal rights, while the coercive power of labour has been undermined and curtailed by (amongst other things) a steady erosion of its legal rights at both the individual and collective levels. These changes in the "structures of mutual coercion" have generated major shifts in the balance of class forces, in the distribution of the social product and in the economic dynamics of contemporary capitalism.[132] While they have been depicted as the natural products of economic evolution (or market forces), in reality many of them have been rooted in fundamentally *political* decisions to alter the prevailing structures of rights and power.[133] Hidden behind the free-market rhetoric of irresistible economic imperatives has lurked a ruthless political project designed to restore wealth and power to a few.

The rise of bubblenomics

In certain respects this project has been successful. The emergence of a neoliberal order under the aegis of finance has destroyed the old social compromises and greatly weakened the power of labour. Huge new, amorphous, unorganised, increasingly feminised, easily exploited proletariats have been created, job security reduced and social protection systems weakened, The result of these and the other changes to the "structures of mutual coercion", including those effected through law, has been downward pressure on wages, a marked redistribution of the proceeds of industry from labour to the bondholding class and to small elites in industry and finance,[134] and a sharp reversal of the trend towards greater equality.[135]

In other ways, however, the strategy has failed, doing little to eradicate the structural problems facing contemporary capitalism. The interconnected problems of over-accumulation, overcapacity and underconsumption – what David Harvey calls "the capital surplus absorption problem"[136] – remain unresolved. There remains a serious shortage of outlets for profitable investment in real productive activity. Since the 1980s, stagnation has been gripping the capitalisms of the developed world, with economic performance in the US, Western Europe and Japan steadily deteriorating in terms of standard economic indicators such as gross domestic product (GDP), investment and real wages. Despite the "greatest government sponsored economic stimulus in US peacetime history", the business cycle that began in 2001 and ended in 2007 was the weakest of the post-war period: GDP growth was slow and private sector employment barely shifted.[137] Paradoxically, this was in part because by holding down wages demand was depressed, exacerbating the chronic, systemic tendency towards overcapacity which has been dogging manufacturing industry

131 R L Hale, "Coercion and distribution in a supposedly non-coercive state" (1923) 38 *Political Science Quarterly* 470.

132 See Ireland, "Property", n. 21, above.

133 See R Helleiner, *States and the Re-emergence of Global Finance* (New York: Cornell University Press 1994), arguing that most accounts of the globalisation of financial markets and demise of Bretton Woods downplay the role played by states.

134 As Robert Reich says, the "restructuring" which has gone on within corporations has often entailed "nothing more than redistributing income from employees to shareholders": R Reich, *Harper's Magazine*, May 1996.

135 This trend began in the 1930s and continued to the end of the 1970s. By the turn of the millennium, however, inequality was returning to the levels of the pre-1914 period.

136 Harvey, *The Enigma*, n. 3 above, p. 45

137 Brenner, "What is good", n. 130 above.

worldwide.[138] It is in the attempts to overcome these problems that we can locate some of the roots of financialisation and of the financial crises which have afflicted capitalism with increasing regularity.

Since the beginning of the long downturn, governments have regularly responded to these problems by fostering debt-financed demand. Initially, traditional Keynesianism and public borrowing were deployed, but from the 1990s, beginning in the US but spreading elsewhere, governments moved towards more balanced budgets and began to use private borrowing by households and corporations as alternative sources of economic stimuli. At the same time, asset bubbles were nurtured to suck up surplus capital by creating new investment and profit-making opportunities in financial activity. The "wealth effect" of these bubbles made possible still further borrowing by corporations and households benefiting from increased paper wealth. In the late 1990s, this "asset-price Keynesianism" centred on the explosion in equity prices associated with the dot.com bubble and the alleged rise of a "New Economy". When this particular bubble burst, a new one was nurtured, centred on housing and leveraged lending. Interest rates were kept low to encourage borrowing, and in the US in particular the authorities pressed credit card companies to develop colour- and gender-blind risk models which created equal opportunities to become indebted, turning a blind eye to increasingly lax lending practices and mortgage lending standards. This pseudo-democratisation of credit, debt and (share and house) ownership, which emerged as a sort of alternative social welfare programme, was portrayed by governments as a positive step in the direction of ownership societies.

In many places, of course, the result was rapid rises in the price of residential property, accelerated in the US by the sub-prime super-booster. With annual house price increases coming to be seen as normal, households were prompted to engage in yet more borrowing on the back of the bubble-inflated value of their homes. From the mid-1990s private debt relative to national income rose rapidly.[139] This not only helped to keep many economies afloat but, coupled with the intensification of labour and increase in two-earner households, enabled workers to maintain or even raise their living standards despite increasingly stagnant real wages. It also created a false sense of prosperity and helped to build positive political support for the ruling neoliberal ideas of the age. Financial markets and increasing debt thus acted as both mechanisms for disciplining workers and mechanisms for social integration.[140] In addition to contributing to the titanic growth of multi-layered debt, the residential property bubble also underpinned new financial bubbles, for the transformation of the prospective income streams created by debt – whether from mortgages or credit cards or student loans – into financial property ("securitisation") acted as a basis for further extensions of credit. Sub-prime mortgages could be issued (originated), bundled together (securitised) and sold on (distributed). Moreover, the risks associated with them could, it seemed, not only be spread but engineered away through the use of financial innovations such as credit default swaps (CDSs). For the financial institutions concerned, it was

138 It is because of the conflict between the need for lower real wages to restore profitability and the need for higher wages to restore demand (coupled with the problem of over-indebted households) that some are pessimistic about the prospects of a rapid, market-based recovery: see dos Santos, "At the heart", n. 112 above.

139 J B Foster and F Magdoff, "Financial implosion and stagnation"(November 2008) *Monthly Review* 6–7; see also dos Santos, "At the heart", n. 112 above.

140 "Constrained in what they could get from their labour", Panitch et al. observe, "US workers were drawn into the logic of asset inflation . . . not only via the institutional investment of their pensions, but also via the one major asset they held (or could aspire to hold) . . . their family home". In the US, "high levels of consumption were sustained by the accumulation of household debt and the intensification of family labour (more family members working longer hours under more severe conditions subject to the growing discipline of having to meet debt payments": Panitch et al., "Political economy", n. 113 above, p. 264.

extraordinarily lucrative business and from the late 1990s their profits headed off into the stratosphere.[141] At the same time more and more hitherto non-financial corporations became quasi-financial entities. In recent decades, in the absence of sufficient outlets for profitable investment in real productive activity, "profit-making [has come to] occur increasingly through financial channels rather than through trade and commodity production".[142] Pension privatisation, by adding to the growing volume of capital seeking outlets, played a part in fuelling these processes.

For institutional investors, operating in an increasingly competitive environment, these new forms of financial property were manna from heaven. Faced by modest rates of return on relatively low-risk financial property forms and desperately searching for higher yields, institutions began to stretch the boundaries of risk and speculation ever further, pouring money into exotic financial instruments and securities like those backed by American sub-prime mortgages. Indeed, in his testimony before the Angelides Commission, Alan Greenspan suggested that the crisis was as much a product of greedy investors (especially in Europe) searching for high-yielding bonds as it was of the misguided policies of the Fed, greedy bankers or fraudulent American mortgage lenders.[143] Even allowing for the blame-shifting here, there is no doubt that the near-insatiable appetite of investors for these securities made it possible for banks to lend to ever-poorer borrowers. We now know, of course, that the banks held on to a large number of them themselves, often via the rapidly rising "shadow banking system", so that when the crash came they were on the books of institutions across the world. As dos Santos observes, "the very idea that defaults on home mortgages would have wiped out leading international financial institutions, and triggered a worldwide recession, would [previously] have been unthinkable",[144] but not only had everyday life been increasingly "financialised", Wall Street had become "more and more dependent on the mundane world of US mortgage and consumer debt": the worlds of high and low finance, of finance and poverty had become entwined.[145] The supporters of pension privatisation were thus quite right when they argued that it would "spur financial market development by creating demand for new financial instruments and institutions",[146] but quite wrong to think that this financial innovation would promote "market completion" and a superior, more perfect form of free-market capitalism. As Lord Turner recently observed, much of the financial innovation was "of little value" and "socially useless".[147]

The end of neoliberal hegemony?

The US housing bubble deflated when mortgage interest rates increased and marginal borrowers, "pressed against the limits of continually increasing working hours",[148] found themselves unable to meet their debt repayments. As foreclosures rose, the income streams which gave value to these mortgage-backed securities evaporated, as did the balance sheets of many financial institutions. It was no longer possible to conceal the fact that the ballooning and increasingly speculative financial superstructure had seriously outgrown the

141 See Foster and Magdoff, "Financial implosion", n. 139 above, p. 7.

142 G Krippner, "The financialization of the American economy" (2005) 3 *Socio-Economic Review* 173.

143 "Blame Europe, former Federal Reserve boss tells US enquiry into financial crisis", *The Guardian,* 7 April 2010. Europe means here primarily Germany. The commission, which was otherwise known as the Financial Crisis Inquiry Commission, was set up by the US government to investigate the causes of the financial crisis.

144 dos Santos, "At the heart", n. 112 above, p. 4.

145 See Panitch et al., "Political economy", n. 113 above, p. 253.

146 James, "New systems", n. 16 above, p. 7.

147 Lord Turner, interview with *Prospect* magazine, 27 August 2009; see also P T Larsen, "Bank regulation needs straightening out", *Financial Times*, 30 March 2009.

148 Panitch et al., "Political economy", n. 113 above, p. 270.

productive base. The crisis was, however, only the latest in a series, from the third-world debt crisis of the 1970s and 1980s, to the US savings and loans scandal, to the "historic misallocation of capital" of the dot.com bubble.[149] The new financialised, neoliberal capitalism has since its inception been characterised by unpredictable currency fluctuations, reckless capital movements, asset bubbles, growing financial instability and regular financial crises. It is the sheer scale of the current crisis that is unprecedented and it was this that prompted some to herald it as marking the beginning of the end for neoliberalism.

It is worth remembering, however, that the East Asian crisis of 1997–98 also prompted talk about the end of neoliberal hegemony and need to build a "new international financial architecture". In the event, however, it was contained in the periphery and attributed to faults in the countries affected, in particular their "crony capitalisms". Instead of fundamental change, we saw the development of global standards of "good governance" aimed principally at creating a sounder platform for further financialisation, holding out to compliant developing countries the promise of foreign capital and to Western investors the promise of new, secure investment outlets. By contrast, the current crisis struck at the heartlands of global capitalism, initially propelling everyone from finance ministers to central bankers to investors into a state of dazed confusion.[150] Since then, however, the ruling elites have regained their bearings and set purposively about patching the system up. Indeed, the responses of governments have laid bare not only the sheer economic and political power of finance but the degree to which neoliberal ideas, policies and practices have become institutionally and culturally embedded.

In the corporate governance context, for example, while there have been mutterings about executive pay and the "dumb" idea of shareholder value,[151] there are a few signs that the financialised, Anglo-American, shareholder-oriented model of the corporation is going to be abandoned. The OECD, for example, concedes that the crisis has highlighted various corporate governance weaknesses, but argues that there is "no urgent need" to revise its Principles, merely a need to ensure their better implementation.[152] The OECD has also made it clear that, although the stock market bounce of 2009 recouped less than half of the investment losses suffered in 2008,[153] it does not intend to abandon pension privatisation. The crisis may have "severely dented the confidence of investors in many countries in DC [defined contribution] schemes" and caused some to "retreat from pension privatisation", but the OECD's firm recommendation is that governments "stay the course".[154] Indeed, the organisation's great fear is that some countries, especially in Central and Eastern Europe, might follow the Argentinian example and turn back towards PAYG public pension schemes when the need is for *more* not less to be invested in private pensions to close the

149 Brenner, "What is good", n. 130 above, p. 29.

150 "Our world is broken – and I honestly don't know what is going to replace it", declared Bernie Sucher, head of Merrill Lynch's operations in Moscow. "The compass by which we steered as Americans has gone.": G Tett, "Lost through creative destruction", *Financial Times*, 10 March 2009.

151 In an interview with the *Financial Times*, reported on 12 March 2009, Jack Welch, the celebrated former head of General Electric stated that "on the face of it, shareholder value is the dumbest idea in the world". This view was endorsed by Lambert, "Does business have a role?", n. 2 above.

152 OECD, Steering Group on Corporate Governance, "Corporate governance and the financial crisis", 24 February 2010.

153 OECD (July 2010) *Pension Markets in Focus*, issue 7. Pension fund assets, it reports, were "struggling" to return to pre-crisis levels. Moreover, "new challenges" were appearing: the onset of retirement of the baby-boom generation, uncertainty over the strength of the economic recovery, the weakness of public bond markets and possible regulatory changes.

154 Antolin and Stewart, "Private pensions", n. 100 above, pp. 2, 5.

"retirement savings gap".[155] Governments need to "expand coverage among middle income workers and low earners" and to ensure that "the general public . . . [is] better informed on the virtues and challenges of their pension system". Investors need to "become more conscious of the risks they face" and to adopt more "suitable" strategies.[156] Quite what this means is unclear. On the one hand, some are admonished for incurring large losses by putting too many eggs into the equity basket (presumably in an attempt to close the retirement savings gap), while others are admonished for being "excessively conservative" and engaging in "too little equity and foreign investment".[157] The OECD, it seems, thinks people need to learn to gamble more skilfully. The UK is also sticking with pensions privatisation and is currently trying to reinvigorate pension saving by compelling employers to automatically enrol employees into workplace pension schemes or the new National Employment Savings Trust (NEST) and to make 3 per cent contributions. Introducing private pensions to low and middle income earners, argues Pensions Minister, Steve Webb, constitutes a "social revolution".[157a] The crisis does not, then, appear as yet to have significantly eroded belief in the key tenets of neoliberal ideology, at least amongst policymakers in the advanced capitalist economies of North America and Europe.[158] Is there nevertheless, as some have suggested, a legitimation crisis looming?

Support for, or acquiescence in, the changes associated with the neoliberal revolution has been garnered in many ways. The idea that it is pointless trying to resist the irresistible economic logic of the market has been relentlessly promoted. Rather than futilely standing in the path of the juggernaut of history both individuals and countries have been urged to accept the imperatives of the market, to adapt, and to make themselves more competitive. The idea that the economic logic of the market is beneficent, encouraging individual self-reliance and responsibility, creating new opportunities for the exercise of individual freedom, and operating to ensure that resources are efficiently allocated and aggregate wealth and welfare maximised, has also been endlessly promoted. As we have seen, the popular attraction of this curious mix of choice and no choice – of compulsion combined with freedom and aspiration – has been strengthened by the propagation of a vision of increasingly inclusive, free, wealthy, market-based ownership societies in which everyone either is, or can reasonably aspire to be, a financial property- and home-owning member of the middle class. Widening home- and share-ownership have been presented as democratising initiatives which create bonds between divergent interests; as homogenising and harmonising measures which eradicate class difference.[159] In this context, investor protection has been presented as something which is both necessary and of benefit to the social whole and not merely the bondholding class.

This vision has unravelled. The financial crisis has not only exposed the chimerical nature of the attempt to create ownership societies populated by worker capitalists, it has

155 Prior to the onset of the crisis, they argue, some countries were already failing to accumulate a sufficiently large asset pool to cover the retirement savings "gap" caused by falling public pension benefits.

156 OECD, *Private Pensions Outlook 2008* (Paris: OECD 2008) pp. 1–4.

157 Thus, "pension fund risk management needs to be strengthened to reduce exposure to unduly risky investments", while the conservative post-crisis "shift in asset-allocation patterns" is identified as a particular problem, for risking "locking in portfolio losses and could reduce the potential of funds to generate retirement incomes in future": Antolin and Stewart, "Private pensions", n. 100 above, pp. 2–4.

157a See S Webb, "Pensions revolution is a social landmark", *The Independent*, Raconteur section, 28 February 2011.

158 The story is rather different in other parts of the world such as South America.

159 Soederberg, *Corporate Power*, n. 60 above, pp. 27, 90–110.

brought the "golden age of home ownership" to an end.[160] If, as Fiona Haines suggests, the promise of future prosperity is, ideologically, often more important than its fulfilment, not least by rendering present deprivations palatable, this unravelling might prove significant.[161] Much will turn, of course, on how people experience and interpret the age of austerity. Will they simply accept it and adapt to falling living standards, economic stagnation, uncertain futures, and poorer and later retirements, particularly if financial institutions, corporate executives and other financial intermediaries continue to reap large and disproportionate rewards? Will it be possible to convince people that (neoliberal) capitalism is good for them and not just a small, wealthy elite, no matter how stark the contrast between the socialisation of risks and losses and the continuing privatisation of profits? Is it going to be possible to construct a new, positive vision of future prosperity?[162]

Rethinking regulatory reform

At the moment, all that seems to be on offer is a vision of unevenly shared sacrifice driven by the irresistible imperatives of "the markets".[163] The coalition government does not deny that policy mistakes were made – on the contrary, the profligacy of the previous administration have been emphasised, as has the need for regulatory reform – but argues that there is now little or no room for manoeuvre. The path of future policy is to a significant extent dictated by economic imperatives – and in particular, of course, by the demands of financial markets and bondholders.

This economically determinist view urgently needs to be challenged, for it is stifling consideration of the enormous range of policy options and institutional arrangements which are available. We need to dispel the idea that there exists a universal, apolitical, purely economic rationality (that of the market) which cannot be resisted and the idea that the goal of policy should be regulatory reforms aimed at correcting "market failures" and at creating perfectly functioning markets. We need instead to recognise that the allegedly inexorable economic forces before which we are being asked to bow are themselves in significant part legal and political constructs. Legal scholars are unusually well placed to do this, for, as noted earlier, they are, or should be, aware that "there is really no such thing as a free market";[164] that "a market cannot be defined except with reference to the specific rights/obligations structure that underpins it, and that since these rights and obligations are determined through a political process . . . all markets have a fundamentally political origin".[165] Markets are legal, political (and, therefore, regulatory) *products*, not spontaneously arising, pre-regulatory, pre-legal and pre-political phenomena which are merely *subjected* to regulation. They are or should also be aware that the property rights underpinning market

160 In commenting on their report, "Widening the rental housing market" (August 2010), Sarah Webb, Chief Executive of the Chartered Institute of Housing said that "a golden age of home ownership" was "coming to an end".

161 F Haines, "Socializing economic relationships in the context of the financial and climate crises: what can regulation offer?" (paper delivered to Socializing Economic Relationships: New perspectives and methods for analysing transnational risk regulation, workshop held in Oñati, Spain, June 2010).

162 The political importance of such a vision was implicitly recognised by Ed Miliband when he sought to depict the Labour Party as the party of "optimism" in his conference speech in September 2010.

163 An important distinction is to be drawn, Haines argues, between aspirational sacrifices and imposed sacrifices, "Socializing", n. 161 above.

164 H-J Chang, *23 Things They Don't Tell You about Capitalism* (London: Allen Lane 2010), p. xvi.

165 H-J Chang, "An institutionalist perspective on the role of the state: towards an institutionalist political economy" in Chang, *Globalisation*, n. 25 above, pp. 98–9.

exchange are themselves regulatory products, contingent bundles of rights which can, have been, and are legally constituted in many different ways.[166]

It also means recognising that all property rights, including financial property rights, are sources of power and that the precise content of these rights bundles – their political/legal/regulatory constitution – is, therefore, vitally important, a key determinant of the power their holders can exercise. Markets are not neutral spaces for social interaction, but sites in which often very unevenly distributed power is exercised.[167] Even in a so-called free-market society, what appear to be (and are presented as) the objective outcomes of impersonal markets are, in fact, significantly shaped by certain (explicit and implicit) political and legal decisions about property rights and other legal entitlements.[168] These decisions shape the different rationalities by which different markets operate and are crucial determinants of the distribution of the social product. From this perspective, the growing power of financial interests which underlies the neoliberal revolution is in significant part a product of enhancements to the bundles of rights possessed by financial property owners, not least their right to move freely around the world (financial liberalisation) and to claim protection from state interventions that diminish their value, even if democratically mandated (the new constitutionalism). In an extreme example of "regulatory capture", the bondholding class and various financial elites have established favourable new rights–obligations structures and then sought to create level playing fields ("free" markets) in which those rights can be coercively exercised. They have "shaped the environment for [their] own convenience", and in doing so laid claim to a growing share of the social product.[169] Aided by the notion that money makes more money by itself, at no cost to anyone, policies which benefit a few have been passed off as both inevitable and as socially beneficial. In seeking to account for many of the legal and other developments associated with the neoliberal revolution (including the pre-occupation with investor protection), therefore, we need to *move* power from the margins to the centre of the stage. These developments and, indeed, the bail-outs are much better understood when viewed through a prism of (class) power than through a neoliberal prism of "deregulation", "efficiency" and "market imperatives". In reality, far from marking a hallowed path to growth and development, financial liberalisation has simply paved the way for new forms of exploitation, imperialism and socio-economic inequality.

While the current thirst for regulatory reform should, therefore, be welcomed, we need radically to rethink what it entails. At present the term "regulation" tends to be used in a way which assumes that both property rights and markets are spontaneously arising, natural products which somehow pre-exist regulatory intervention. Regulation is portrayed as

166 See, for example, A M Honore "Ownership" in A G Guest (ed.), *Oxford Essays in Jurisprudence* (Oxford: Clarendon 1961); see also Ireland, "Property", n. 21 above. This is unusually clear in respect of the intangible rights to receive future revenues (the financial property forms) that lie at the heart of contemporary capitalism and the current crisis. With financial property, there is no concrete object of property independent of law to which the rights relate; the very "thing" that is owned is a legal construct, even if once constituted it tends to take on a reified life of its own. These property forms are regulation all the way down.

167 Soederberg, *Corporate Power*, n. 60 above, p. 28.

168 See Chang, "An institutionalist perspective", n. 165, pp. 94–5: "The establishment and distribution of property rights and other entitlements that define the endowments that neoclassical economics take as given", Chang observes, "is a highly political exercise", from which he concludes that "all prices are political".

169 Canterbery, *Wall Street*, n. 67 above, p. 5. Indeed, in constructing these new forms of financial power Western law has performed the miracle of enabling the "rich, developed, and often ex-colonial states . . . to . . . continue extracting wealth from the poorest countries" in an era of apparent de-colonialisation. Gun-boat diplomacy and costly imperial administrations have largely (though not entirely) been replaced by other, less militaristic, less politically overt, purely "economic" modes of extraction: see W Mansell, "Legal aspects of international debt" (1991) 18 *Journal of Law & Society* 381.

something which is imposed on them from the outside, rather than as something intrinsic to them, something which actually constitutes them. As a result, many regulations are not seen or acknowledged as such, but treated as natural, as having a pre-regulatory existence, and removed from the arena of regulatory reform.[170] If we are to tackle the problems we now face, these invisible regulations need to be identified and opened up for debate. It should be recognised that states determine the bundles of rights held by financial property owners and the terms on which international financial markets operate, and the idea of regulatory reform should be widened to encompass consideration of them. Rather than being treated as unchangeable and unchallengeable natural givens, these bundles of rights need openly to be recognised as social artefacts, subjected to critical analysis and, if necessary, to reform. Regulatory reform should not concern itself only with investor protection and creating a more secure environment for financialisation, but with consideration of the proper ambit and scope of investors' rights and, indeed, obligations. The present (socially constructed) rights–obligations structures systematically favour the interests of the bondholding class to the disadvantage of everyone else.[171]

There is a precedent, for this is, in effect, what happened at Bretton Woods. Both John Maynard Keynes and Harry Dexter White, the principal architects of the system, wanted to secure for governments the right to control their domestic capital markets and to prevent international capital from "operat[ing] against what the government deemed to be the interests of any country". They wanted, in particular, to protect newly emerging welfare states from capital flights induced by a desire to evade the "burdens of social legislation", and to promote "genuine new investments for developing the world's resources".[172] Bretton Woods thus promoted the free trade of goods within a restrictive and decidedly *il*liberal set of international financial arrangements characterised by fixed exchange rates and extensive capital controls. In restricting and regulating the international movement of capital, both Keynes and White acknowledged that they were limiting the bundles of rights possessed by financial property owners and thus significantly diminishing their market power.[173] As White observed, the new system would mean "less freedom for owners of liquid capital" and place "restriction[s] on the property rights of the 5 or 10 percent of persons in foreign countries who have enough wealth or income to keep or invest some of it abroad".[174] In similar vein, Keynes recognised that the implementation of the accord would precipitate "keen political discussions [about] the position of the wealthier classes

170 See Chang, *23 Things*, n. 164 above, pp. 1–10.

171 As dos Santos observes, amongst other things they permit lenders to realise "high effective interests rates", underpinning the sustained high profitability enjoyed by financial intermediaries lending to ordinary consumers, "At the heart", n. 112 above, p. 6.

172 E Johnson and D Moggridge (eds), *The Collected Writings of John Maynard Keynes* vol. 25 (London: Macmillan 1980), p. 17. Keynes favoured "legitimate capital movements". According to Helleiner, Bertil Ohlin summarised the views of many international economists in 1936: "There is a decisive difference between the role of such transfers [capital movements] and the functions of an exchange of commodities. The latter is a prerequisite of prosperity and economic growth, the former is not.": Helleiner, *States*, n. 133 above, pp. 33–8.

173 The goal, in the words of Henry Morgenthau, American Treasury Department Secretary, was to "drive the moneylenders from the temple of international finance": Helleiner, *States*, n. 133 above, p. 4.

174 H D White, "Preliminary draft proposal for a UN stablization fund (April 1942)", in J Keith Horsfield, *The International Monetary Fund 1945–65* vol. 3: Documents (Washington: IMF 1969), pp. 66–7.

and the treatment of private property".[175] Both believed that the control of capital movements was a prerequisite of effective domestic economic and social management.

The desire to make finance servant rather than master[176] lies behind not only the calls for the establishment of a Bretton Woods II, but many of the other reform proposals which have surfaced in recent months – for the paring down of shareholder rights (introducing minimum ownership periods before shares accrue voting rights, for example),[177] for the imposition of greater restrictions on takeovers, for the introduction of maximum executive–worker remuneration ratios, for Tobin-style taxes, and so on. Any proposals which diminish the rights and privileges of the bondholding class will, of course, be highly controversial and fiercely contested. As J A Hobson observed during the depression of the 1930s, any political changes which entail changes to property rights tend to be portrayed by property owners as outrageous, wicked and wrong, infringements of natural rights and freedoms. It is, he suggested, for this reason that academics and policymakers are reluctant to subject property and property rights to the rational scrutiny to which they subject other social institutions. The "refusal to apply clear reasoning to unveil the defects of political and economic institutions, and the . . . class distinctions associated with them", Hobson argued, "is in some measure due to the stubborn objection of rationalists to apply to property, income, profit, and other economic concepts the same relentless logic they apply to [other] concepts".[178] Given the seriousness of the situation in which we currently find ourselves, it is precisely clear reasoning of this sort that we need.

Even before the current debacle, belief in the virtues of neoliberalism (and pension privatisation) persisted not because of the results it had delivered but because of the pressure exerted in its favour by financial elites, their representatives in powerful national and international agencies, and academics and policymakers seduced by simplistic theories about the benefits of so-called free markets and free financial markets in particular.[179] The crisis presents an opportunity to challenge not only the intellectual hegemony of these ideas and their false claims about the nature of markets, but the debilitating economic determinism and distorted and obfuscatory efficiency-based explanatory and normative frameworks which have underlain so much recent legal scholarship. It presents us, in other words, with an opportunity to re-orient and re-enrich that scholarship, to reconnect it with the vital and complex issues of productive efficiency, justice, ethics and human well-being with which law and regulation should be concerned, and to recognise that the range of institutional possibility is far greater than we have been led to believe.

175 Letter to Roy Harrod, 19 April 1942, Johnson and Moggridge (eds), *Collected Writings*, n. 172 above, pp. 35, 149. Keynes was not opposed to finance per se, differentiating the active financier who sensed opportunities for investment in particular productive sectors from the parasitic coupon-clipper (whom he described as an investor without a function) living off interest payments and dividends. For the latter he advocated euthanasia, slow death through declining interest rates.

176 In Lawrence Kraus' words, Bretton Woods assigned to finance a kind of "second class status", see Helleiner, *States*, n. 133 above, p. 5.

177 In the speech cited earlier ("Does business have a role?", n. 2 above), Richard Lambert cited with approval Dave Packard's suggestion that shareholders should settle for an "adequate profit".

178 J A Hobson, *Confessions of an Economic Heretic* (London: Allen & Unwin 1938).

179 On this, see Charles Ferguson's revealing film, *Inside Job* (2010).

NILQ 62(1): 33–55

Gordian knots in Europeanised private law: unfair terms, bank charges and political compromises

James P Devenney*

Durham University

Introduction

In recent years there has been a significant amount of public concern in the United Kingdom about the charges levied by banks on personal account holders in respect of unauthorised overdrafts (and similar charges[1]).[2] Indeed, it would appear that many of these account holders made complaints about such charges to the Office of Fair Trading (OFT).[3] Moreover it seems that "many thousands"[4] of these account holders have challenged such charges on the ground, inter alia, that they are unfair for the purposes of the Unfair Terms in Consumer Contracts Regulations 1999 (the Regulations),[5] the regulations which, as is well known, seek to transpose the (minimum harmonisation) EC Council Directive 93/13/EEC on Unfair Terms in Consumer Contracts (the Directive).[6]

In early 2007 the OFT commenced an investigation into such charges.[7] It quickly transpired that a key issue related to whether or not Regulation 6(2) circumscribed any claim that the charges were unfair for the purposes of the Regulations. Regulation 6(2) owes its existence to Article 4(2) of the Directive. Article 4(2) provides:

> Assessment of the unfair nature of the terms shall relate neither to the definition of the main subject matter of the contract nor to the adequacy of the price and remuneration, on the one hand, as against the services or goods supplies [*sic*] in exchange, on the other, in so far as these terms are in plain intelligible language.

* Director, Durham University Institute of Commercial and Corporate Law. I would like to thank Dr Mel Kenny, Claire Devenney, Richard Coleman, Dr Amandine Garde, Dr Warren Swain and Karen Fairweather for their assistance. The usual caveats apply.

1 A useful summary of which can be found in the OFT's joint reply and defence to the counterclaims (dated 11 November 2007) in *OFT* v *Abbey National plc* [2008] EWHC 875 (Comm), which is available at www.oft.gov.uk/shared_oft/personal-current-accounts/OFT's-joint-reply-and-def.pdf.

2 See, for example, OFT, *Personal Current Accounts in the UK* (OFT 918, April 2007), p. 2, which was followed by, inter alia, OFT, *Personal Current Accounts in the UK: A market study* (OFT 1005, July 2008).

3 Ibid.

4 *Office of Fair Trading* v *Abbey National plc* [2009] UKSC 6, para. 17, per Lord Walker.

5 See, for example, the narrative in *OFT* v *Abbey National plc* [2008] EWHC 875 (Comm), para. 2, per Andrew Smith J.

6 (1993) OJ L95/29. In relation to the consequences of the Directive being one of minimum harmonisation, see *Caja de Ahorros y Monte de Piedad de Madrid* v *Asociación de Usuarios de Servicios Bancarios* C-484/08, [2010] 3 CMLR 43.

7 See, for example, www.oft.gov.uk/news-and-updates/press/2007/67-07.

Ultimately, the OFT issued proceedings, against (and with the agreement of) various banks and building societies (the banks), centring on the correct interpretation of Regulation 6(2) and its contention that Regulation 6(2) did not circumscribe any claim that the relevant charges were unfair for the purposes of the Regulations.[8] The issue was, for a period of time, considerably widened by the banks' counterclaim to include, for example, issues relating to whether or not the common law rule against penalties was engaged; yet, ultimately, the case focused on the Regulations. At first instance,[9] and in the Court of Appeal,[10] it was held (albeit for different reasons) that Regulation 6(2) did not prevent the relevant charges from being characterised as unfair under the Regulations.

Nevertheless a further appeal by the banks was subsequently allowed by the Supreme Court,[11] the immediate aftermath of which was that the OFT decided not to pursue its investigation into such terms under the Regulations.[12] This paper seeks to analyse the decision of the Supreme Court in *OFT* v *Abbey National plc*, and its implications for the Europeanisation of private law. More specifically this paper presents the decision of the Supreme Court in *Abbey* as another example of the uneven interpretation of the Regulations, and by implication the Directive, in the United Kingdom. Yet there is a wider issue in *Abbey*, viz. the potential impact of political compromises on EU harmonisation agendas; in particular it is argued that Article 4(2) is, essentially, an uneasy and somewhat opaque political compromise provision, the knot of which is only exacerbated by the backdrop of the (almost) chameleonic rationalisations of EU legislation in the consumer arena. Indeed, even leaving aside the evolving nature of internal market considerations,[13] the courts of the Member States are only provided with, at best, vague co-ordinates on how to interpret Article 4(2). In such circumstances, a robust and efficient reference process to the European Court of Justice, to act as a compass, seems essential. Yet *Abbey* reveals a real reluctance on the part of the Supreme Court to refer this issue to the European Court of Justice, and this paper also explores some of the possible reasons behind such a stance.

The Unfair Terms in Consumer Contracts Regulations 1999

At the outset, it will be helpful to briefly outline the Unfair Terms in Consumer Contract Regulations 1999. As noted above, these Regulations seek to transpose the (minimum harmonisation) EC Council Directive on Unfair Terms in Consumer Contracts.[14] In general

8 *OFT* v *Abbey National plc* [2009] UKSC 6, para. 18, per Lord Walker.

9 [2008] EWHC 875 (Comm).

10 [2009] EWCA Civ 116.

11 [2009] UKSC 6.

12 See www.oft.gov.uk/news-and-updates/press/2009/144-09 where the following is stated: "After detailed consideration of the judgment and of the various options available to it, the OFT has concluded that any investigation it were to continue into the fairness of current unarranged overdraft charging terms under the UTCCRs [the Regulations] would have a very limited scope and low prospects of success. Given this, it has decided against taking forward such an investigation. The OFT nevertheless continues to have significant concerns about the operation of the market for personal current accounts. Despite some recent and planned improvements by banks, particularly around transparency and customer switching, it believes fundamental changes are still required for the market to work in the best interests of bank customers. Banks earn around a third of their personal current account revenues from unarranged overdraft charges that are difficult to understand, not transparent and not subject to effective consumer control."

13 Cf. J H H Weiler, "The transformation of Europe" (1991) 100 *Yale Law Journal* 2043; L W Gormley, "Competition and free movement: is the internal market the same as the Common Market?" (2002) *EBL Rev* 522.

14 93/13/EEC.

terms, these freestanding Regulations follow closely the wording of the Directive.[15] Regulation 4 states that the Regulations deal with "unfair terms in contracts concluded between a seller or a supplier and a consumer"; and Regulation 8 states that an "unfair term" is not binding on the consumer. A seller/supplier is defined as "any natural or legal person who . . . is acting for purposes relating to his trade, business or profession" whereas a consumer is defined as "any natural person who . . . is acting for purposes which are outside his trade, business or profession".[16] Regulation 5(1) (which is supplemented by an indicative but non-exhaustive list of the terms which might be regarded as unfair)[17] states that:

> A contractual term which has not been individually negotiated shall be regarded as unfair if, contrary to the requirement of good faith, it causes a significant imbalance in the parties' rights and obligations arising under the contract, to the detriment of the consumer.

The test of "unfairness" in Regulation 5 is circumscribed by other parts of the Regulations. In particular, Regulation 6(2) provides:

> In so far as it is in plain intelligible language, the assessment of fairness of a term shall not relate—
>
> (a) to the definition of the main subject matter of the contract, or
>
> (b) to the adequacy of the price or remuneration, as against the goods or services supplied in exchange.

It was Regulation 6(2) which was at the centre of the bank charges litigation.

The bank charges litigation

The OFT is empowered to make collective challenges to unfair terms under the Regulations, as is envisaged by Article 7 of the Directive.[18] Article 7 of the Directive provides:

> Member States shall ensure that, in the interests of consumers and of competitors, adequate and effective means exist to prevent the continued use of unfair terms in contracts concluded with consumers by sellers or suppliers . . . The means . . . shall include provisions whereby persons or organizations, having a legitimate interest under national law in protecting consumers, may take action . . . for a decision as to whether contractual terms drawn up for general use are unfair, so that they can apply appropriate and effective means to prevent the continued use of such terms.

As noted above, in 2007 the OFT issued proceedings, against (and with the agreement of) various banks and building societies, centring on the correct interpretation of Regulation 6(2) and its contention that Regulation 6(2) did not circumscribe any claim that the relevant charges were unfair for the purposes of the Regulations.[19] At first instance, Andrew Smith J held that the relevant terms were not protected by Regulation 6(2) from being characterised as unfair under the Regulations; in particular the learned judge, focusing on the wording of Regulation 6(2), felt that the charges were not paid *in exchange* for any

15 The Explanatory Notes to the Regulations state: "These Regulations revoke and replace the Unfair Terms in Consumer Contracts Regulations 1994 (S.I. 1994/3159) which came into force on 1st July 1995. Those Regulations implemented Council Directive 93/13/EEC on unfair terms in consumer contracts (O.J. No. L95, 21.4.93, p. 29). Regulations 3 to 9 of these Regulations re-enact regulations 2 to 7 of the 1994 Regulations with modifications to reflect more closely the wording of the Directive."

16 Regulation 3.

17 Schedule 2.

18 See also Enterprise Act 2002, Part 1.

19 *OFT* v *Abbey National plc* [2009] UKSC 6, para. 18, per Lord Walker.

services.[20] Such a stance is particularly attractive in relation to so-called "unpaid item charges" where a bank *refuses to* honour an instruction (and thereby provide a "service") on an account with insufficient funds for the instruction.[21]

The Court of Appeal[22] dismissed the appeal[23] although its reasoning did not entirely match the reasoning of the learned judge. Essentially, the Court of Appeal, taking its lead from *Director General of Fair Trading* v *First National Bank plc*,[24] made a distinction between core and ancillary terms, only the former of which come within Regulation 6(2);[25] and held that the charges in question were ancillary terms and, hence, not covered by Regulation 6(2).[26] Nevertheless a further appeal by the banks was subsequently allowed by the Supreme Court.

Regulation 6(2)

Before considering the judgment of the Supreme Court in *OFT* v *Abbey National plc*, it is helpful to identify a number of issues which arise in relation to Regulation 6(2). The first issue which arises in relation to Regulation 6(2) concerns the rationale behind Article 4(2)[27] of the Directive (the Article which Regulation 6(2) seeks to transpose). One of the difficulties here is that the language employed by Article 4(2) and Recital (19)[28] of the Directive is not particularly helpful in this regard; indeed it is, perhaps, unsurprising that this should be so given that the legislative history of Article 4(2) reveals that it is essentially a, not entirely comfortable,[29] compromise provision.[30] Nevertheless, in general terms it seems that Article 4(2) owes its genesis to the distinction between core terms and ancillary terms found in the German Standard Contracts Act 1976;[31] the essential idea being that in a general sense there is only real consent to core terms.[32]

---

20 At paras 406–9 the learned judge stated: "In reality they are not charges in exchange for services involved in making payments, but charges levied because the services are supplied in particular circumstances. While the Banks' evidence shows that additional processes are involved if the customer gives a payment instruction when he does not have funds or a facility to cover it, these are not explained in the documentation provided to customers and, even if the typical customer would suppose that this might be the case, the contract does not identify these additional processes as being provided in exchange for the charges . . . Undoubtedly, as the Banks point out, when a bank makes arrangements with a customer in advance for an overdraft facility, the customer might incur an arrangement fee for the facility as well as interest, but typically the fee for the facility is charged regardless of whether it is used and the interest is charged for the borrowing itself. Clearly the Relevant Charges are not levied in exchange for a facility in that sense."

21 See *OFT* v *Abbey National plc* [2008] EWHC 875 (Comm), para. 403, per Andrew Smith J.

22 Sir Anthony Clarke MR, Lord Justice Waller V-P and Lloyd LJ.

23 [2009] EWCA Civ 116.

24 [2001] UKHL 52.

25 [2009] EWCA Civ 116, para. 69.

26 Ibid. para. 104.

27 Cf. M Chen Wishart, "Transparency and fairness in bank charges" (2010) 126 *Law Quarterly Review* 157.

28 Recital (19) states: "Whereas, for the purposes of this Directive, assessment of unfair character shall not be made of terms which describe the main subject matter of the contract nor the quality/price ratio of the goods or services supplied; whereas the main subject matter of the contract and the price/quality ratio may nevertheless be taken into account in assessing the fairness of other terms; whereas it follows, inter alia, that in insurance contracts, the terms which clearly define or circumscribe the insured risk and the insurer's liability shall not be subject to such assessment since these restrictions are taken into account in calculating the premium paid by the consumer."

29 Cf. also T Hartley, "Five forms of uncertainty in European Community law" (1996) *Cambridge Law Journal* 265.

30 Cf. H-W Micklitz, *The Politics of Judicial Co-Operation in the EU* (Cambridge: CUP 2005), p. 360.

31 See the excellent discussion in C Willett, *Fairness in Consumer Contracts: The Case of Unfair Terms* (Aldershot: Ashgate 2007), pp. 245–53.

32 A de Moor, "Common and civil law conceptions of contract and a European law of contract: the case of the Directive on Unfair Terms in Consumer Contracts" (1995) 3 *European Review of Private Law* 257, p. 268.

The second issue relates to the effect of Regulation 6(2). More specifically, it is debatable whether Regulation 6(2), where applicable, prohibits *any* claim that the term in question is "unfair" under the Regulations, or whether it *only* prohibits a claim that the term in question is "unfair" under the Regulations *on the ground* that it is, in effect, substantively unfair. The former construction might be classified as an "excluded term" approach, whereas the latter construction might be classified as an "excluded assessment" approach.[33] Here again the language employed by Article 4(2) and Recital (19) of the Directive, and Regulation 6(2), is not particularly helpful: Recital (19) arguably tends to the "excluded term" construction whereas Article 4(2) (and consequently Regulation 6(2)) arguably tends to an "excluded assessment" construction. Presumably, however, if the rationale suggested above for Regulation 6(2) is correct, then an "excluded assessment" construction should be adopted on the basis there may not be true consent to terms which were, for example, procured, in effect, by procedural unconscionability.[34]

The third issue relates to the effect of the introductory words of Regulation 6(2) ("[i]n so far as it is in plain intelligible language"): does this mean that if a term which would normally be covered by the exception in Regulation 6(2) is not "in plain intelligible language", the exception (whatever its parameters) provided by Regulation 6(2) ceases to be applicable? Again, presumably, if the rationale suggested above for Regulation 6(2) is correct, then it can be argued that the exception found in Regulation 6(2) does not necessarily apply at all as there may not have been, or at least there is a risk that there will not have been, true consent.[35] Indeed, this is the *result* that the Supreme Court in *Abbey*[36] tended towards when dealing with the problematic case of *OFT* v *Foxtons*.[37]

The fourth issue relates to the applicability of Regulation 6(2). In particular, *if* (as seems to be the case)[38] Regulation 6(2) focuses on particular terms, *which* terms relate to "the price or remuneration"? In a sense, of course, all terms which confer a benefit on the seller/supplier, or impose a burden on the consumer, constitute part of "the price or remuneration". Yet if such an interpretation were adopted in conjunction with an "excluded term" approach, the Directive (and the Regulations) would be almost redundant; and if such an interpretation were adopted in conjunction with an "excluded assessment" approach, the Directive (and the Regulations) would provide little, if any, relief against, so-called, substantive unfairness (which contrasts uncomfortably with some of the terms in the Schedule 2 grey-list).[39] Thus, the important point for this paper is that there must be some means of distinguishing between those parts of "the price or remuneration" (in its widest sense) which attract the attention of Regulation 6(2) and those parts which do not.

The narrow issue before the Supreme Court

The Supreme Court in *Abbey* was at pains to stress[40] that the appeal before it only concerned a relatively narrow issue, viz. whether or not it would be *possible*, at some point, for the OFT to challenge, under the Regulations, the relevant charges on the ground that they were, or might be, excessive in comparison to the services supplied in return. Indeed, Lord Phillips came close to allowing the appeal *partly* on a technical point of pleading: His

33 See *OFT* v *Abbey National plc* [2009] UKSC 6, paras 60–1, per Lord Phillips.

34 Willett, *Fairness in Consumer Contracts*, n. 31 above, pp. 245–53.

35 Ibid. p. 246.

36 See para. 37.

37 [2009] EWHC 1681.

38 E Peel, *Treitel: The Law of Contract* (London: Sweet & Maxwell 2007), p. 296.

39 See, for example, Schedule 2, 1(l).

40 See, for example, *OFT* v *Abbey National plc* [2009] UKSC 6, para. 3, per Lord Walker.

Lordship felt that the charges were not charges for individual services but were part of the banks' overall remuneration for the package of services provided to relevant customers (and, in Lord Phillips' opinion, the OFT had not sought to question the banks' overall charges);[41] and therefore any claim that the charges were excessive, in comparison to the individual services which triggered those charges, was essentially misconceived.[42]

On this basis, the issue of whether or not the relevant charges were unfair for the purposes of the Regulations was not directly at issue in the appeal. Nevertheless, as noted above, the Supreme Court did *seem* to prefer the "excluded assessment" approach to Regulation 6(2)[43] and it did, tantalisingly yet opaquely, suggest that – despite its finding that Regulation 6(2) prevented the OFT from claiming that the relevant charges were excessive in comparison to the services supplied in return – a challenge under the Regulations was still possible.[44] One might, therefore, take issue (notwithstanding any constraints in terms of pleadings) with the narrow approach of the Supreme Court on two grounds.

First, we were told that there were "many thousands"[45] of cases against banks, concerning the issue of whether or not the charges in question were unfair, stayed in the County Courts pending a decision in *Abbey*. Yet the narrow (even technical) approach of the Supreme Court in *Abbey*, combined with the fact that it seemed to leave the door open to challenging the relevant charges under the Regulations, does little for the effective case management of such claims (often brought by litigants in person).[46] As we shall see below, it is, therefore, ironic that the Supreme Court argued – as a reason not to make a reference to the European Court of Justice – that there was a "public interest" in resolving the issue quickly given, for example, the number of stayed cases awaiting its decision.[47]

The second issue concerns the impact of the decision of the European Court of Justice in *Océano Group Editorial SA* v *Murciano Quintero*[48] relating to the issue of whether a national court could unilaterally raise the issue of unfairness (in the sense provided for by the Directive). The European Court of Justice noted:

> As to the question of whether a court seised of a dispute concerning a contract between a seller or supplier and a consumer may determine of its own motion whether a term of the contract is unfair, it should be noted that the system of protection introduced by the Directive is based on the idea that the consumer is in a weak position vis-à-vis the seller or supplier, as regards both his bargaining power and his level of knowledge. This leads to the consumer agreeing to terms drawn up in advance by the seller or supplier without being able to influence the content of the terms . . . *the protection provided for consumers by the Directive entails the national court being able to determine of its own motion whether a term of a contract before it is unfair when making its preliminary assessment as to whether a claim should be allowed to proceed before the national courts.*[49]

41 See O*FT* v *Abbey National plc* [2009] UKSC 6, para. 64 and, particularly, para. 91.

42 See ibid. paras 62–3 and 89–91.

43 See ibid. para. 52, per Lord Walker, paras 61 and 91, per Lord Phillips, and, most clearly, para. 95, per Lord Mance.

44 See ibid. para. 52, per Lord Walker.

45 Ibid. para. 17, per Lord Walker.

46 Ibid.

47 See n. 169 below and text thereto.

48 C-240/98 to C-244/98.

49 Ibid. paras 25–9 (emphasis added).

Océano Group Editorial SA v *Murciano Quintero*, of course, concerned a jurisdiction clause; yet it is not limited to such clauses.[50] The rationale for national courts being able, and indeed it now seems *required*,[51] to unilaterally raise the issue of fairness was set out in the following terms:

> The aim of Article 6 of the Directive, which requires Member States to lay down that unfair terms are not binding on the consumer, would not be achieved if the consumer were himself obliged to raise the unfair nature of such terms. In disputes where the amounts involved are often limited, the lawyers' fees may be higher than the amount at stake, which may deter the consumer from contesting the application of an unfair term . . . It follows that effective protection of the consumer may be attained only if the national court acknowledges that it has power to evaluate terms of this kind of its own motion . . . As the French Government has pointed out, it is hardly conceivable that, in a system requiring the implementation of specific group actions of a preventive nature intended to put a stop to unfair terms detrimental to consumers' interests, a court hearing a dispute on a specific contract containing an unfair term should not be able to set aside application of the relevant term solely because the consumer has not raised the fact that it is unfair. On the contrary, the court's power to determine of its own motion whether a term is unfair must be regarded as constituting a proper means both of achieving the result sought by Article 6 of the Directive, namely, preventing an individual consumer from being bound by an unfair term, and of contributing to achieving the aim of Article 7, since if the court undertakes such an examination, that may act as a deterrent and contribute to preventing unfair terms in contracts concluded between consumers and sellers or suppliers.[52]

In this spirit – and given the case management issue in the background – it is clearly arguable that the Supreme Court should have addressed the question of unfairness in *greater* detail.

Regulation 6(2) in the Supreme Court

Approach to interpretation

The approach which national courts adopt when interpreting legislation which seeks to transpose an EU Directive will impact, of course, on the strength and depth of the EU harmonisation agenda.[53] Accordingly, it is axiomatic that national courts, normally, must adopt an "EU" approach to the interpretation of such implementing legislation: thus, for example, national courts are required to safeguard the effectiveness of the Directive in question[54] and are often required to give concepts used in directives "an autonomous and

50 Cf. *Cofidis SA* v *Fredout* Case C-473/2000.

51 See S Whittaker, "Judicial interventionism and consumer contracts" (2001) 117 *LQR* 215, p. 217. More recently, in *Pannon GSM Zrt* v *Erzsébet Sustikné Györfi*, C-243/08 the European Court of Justice stated (at para. 35): "The reply, therefore, to the second question is that the national court is *required* to examine, of its own motion, the unfairness of a contractual term where it has available to it the legal and factual elements necessary for that task. Where it considers such a term to be unfair, it must not apply it, except if the consumer opposes that non-application. That duty is also incumbent on the national court when it is ascertaining its own territorial jurisdiction." (emphasis added).

52 C-240/98 to C-244/98, paras 26–8.

53 Cf. A Colombi Ciacchi, "Non-legislative harmonisation of private law under the European constitution: the case of unfair suretyships" (2005) 13 *European Review of Private Law* 297.

54 Cf. Case 6/64 *Flaminio* v *ENEL* [1964] ECR 585.

uniform interpretation".[55] Yet, subject to these overriding obligations, as noted below[56] directives *may* provide a unique opportunity to develop national law, and meet EU obligations, in a coherent manner (albeit that this opportunity is not always exploited).[57]

At times the courts of England and Wales have taken, what seems like, a quintessentially EU approach to interpreting the Regulations. For example, in *R (on the application of Khatun)* v *Newham LBC*,[58] where the Court of Appeal essentially had to consider whether or not the Regulations were applicable to land transaction, Laws LJ concluded:

> As for the bite of the various materials I have cited, I consider that the OFT had the better of the argument. First, Mr Underwood's seemingly strong point on the language – that "goods and services" does not include land – is effectively demolished by the impact of the other language texts. "Biens" and its cognates in Italian, Spanish and Portuguese refer to immovables as readily as movables. This alone undercuts a good deal of what Mr Underwood had to say. But more than this: I think, with respect to Mr Underwood, that other aspects of his submissions on this part of the case place an implicit but illegitimate reliance on the large divide in the law of England between real and personal property. He submitted that the Directive should be interpreted as only applying to "contracts for goods and services as an English lawyer would understand those terms". There is plainly no general principle to support such a proposition. Quite the contrary: European legislation has to be read as a single corpus of law binding across the member states. And the proposition leads to absurdity. A licence of land, which transfers no estate, might be covered by the Directive (as the provision of a service), but a lease or tenancy would not. The sale of a fixture, which by English law is treated as part of the land, would be excluded, but the sale of an identical object – say a statue – which was not fixed to the land would be included. In our domestic law these distinctions have a long history and a present utility. In the context of a Europe-wide scheme of consumer protection, they could be nothing but an embarrassing eccentricity.[59]

Yet, overall, the approach to the interpretation of the Regulations (and, therefore, the Directive) by the Courts in England and Wales is a little uneven.[60] This point can be illustrated by reference to the case law in England and Wales on the vexed question of whether or not the Regulations apply to non-professional surety transactions.[61] From a literal point of view, a difficulty with applying the Regulations to such transactions is that

55 See, for example, Case C-287/98 *Luxembourg* v *Linster* [2000] ECR I-6917, para. 43.

56 See text after n. 192 below.

57 Cf. G Teubner, "Legal irritants: good faith in British law or how unifying law ends up in new divergences" (1998) 61 *Modern Law Review* 11.

58 [2004] EWCA Civ 55.

59 Ibid. para. 78.

60 See also, for example, the European Commission's *Report on Directive 93/13/EEC on Unfair Terms in Consumer Contracts* (Com (2000) 248 final), at p. 32, rather optimistically noted that: "An analysis of CLAB [European database on unfair terms in consumer contracts] shows that already 4.4% of the judgments handed down by national courts in the field covered by the Directive refer to the Community text. At the current stage of European construction this is a figure to be proud of and reflects the progressive impact of Community law on the national legal orders." At p. 34 it is noted that: "National courts could have referred many cases to the Court of Justice for a preliminary ruling and it would have been very useful if the judgments of Court of Justice had been able to cast light on the scope of some of the Directive's more obscure provisions. Indeed the doctrine reveals the reluctance of the national courts to refer cases to the Court of Justice in this legal field."

61 See G McCormack, "Protection of surety guarantors in England – prophylactics and procedure" in A Colombi Ciacchi (ed.), *Protection of Non-Professional Sureties in Europe: Formal and Substantive Disparity* (Germany: Nomos 2007), pp. 172–3.

(assuming that the non-professional surety can be classified as a "consumer" for the purposes of the Regulations) it is the non-professional surety *who supplies the service*; whereas the creditor, as beneficiary of the agreement, will usually be acting in the course of business.[62] Therefore, this question is part of the much wider debate as to whether or not, for the purposes of the Regulations, the consumer must be the *recipient of goods or services*.[63]

Support for the view that the Regulations do apply to surety transactions can be found in the Opinion of the European Court of Justice in *Bayerische Hypothetken* v *Dietzinger*.[64] In that case the European Court of Justice had to consider the applicability of Council Directive 85/577/EEC (on contracts negotiated away from business premises) – which applies to particular situations where "a trader supplies goods or services to a consumer"[65] – to surety transactions. In a judgment, which is not without controversy,[66] the European Court of Justice stated that:

> . . . it is apparent from the wording of Article 1 of Directive 85/577 and from the ancillary nature of guarantees that the directive covers only a guarantee ancillary to a contract whereby, in the context of "doorstep selling", a consumer assumes obligations towards the trader with a view to obtaining goods or services from him. Furthermore, since the directive is designed to protect only consumers, a guarantee comes within the scope of the directive only where, in accordance with the first indent of Article 2, the guarantor has entered into a commitment for a purpose which can be regarded as unconnected with his trade or profession.[67]

In reaching this conclusion – which gives a glimpse of how the European Court of Justice might approach this issue in the context of the Directive[68] – the European Court of Justice noted that nothing in the directive required "the person concluding the contract under which goods or services are to be supplied be the person to whom they are supplied"[69] and that surety agreements are merely ancillary to the main contract.[70]

Returning to the jurisprudence in England and Wales, in *Barclays Bank plc* v *Kufner*,[71] Field J – relying heavily on the Opinion of the European Court of Justice in *Bayerische Hypothetken* v *Dietzinger*[72] – held that surety transactions are not excluded from the scope of the Regulations.[73] By contrast in *Bank of Scotland* v *Singh*,[74] Judge Kershaw QC, apparently

62 See also J O'Donovan and J Phillips, *The Modern Contract of Guarantee* (London: Sweet & Maxwell 2003), p. 223.

63 Cf. H G Beale (ed.), *Chitty on Contracts* 30th edn (London: Sweet & Maxwell 2008), para. 15-032.

64 Case C-45/96 [1998] ECR I-1199.

65 Article 1.

66 See M Kenny, "Standing Surety in Europe: Common Core or Tower of Babel" (2007) *Modern Law Review* 175, p. 180.

67 Case C-45/96 [1998] ECR I-1199, para. 20. Although cf. *Berliner Kindl Brauerei AG* v *Andreas Siepert* [2000] ECR 1-1741, paras 25–6 where the European Court of Justice, in considering Council Directive 87/102/EEC for the approximation of the laws, regulations and administrative provisions of the member states concerning consumer credit, noted: "…the scope of the Directive cannot be widened to cover contracts of guarantee solely on the ground that such agreements are ancillary to the principal agreement whose performance they underwrite, since there is no support for such an interpretation in the wording of the Directive … or in its scheme and aims".

68 See. Beale (ed.), *Chitty on Contracts*, n. 63 above, paras 44–139.

69 Case C-45/96 [1998] ECR I-1199, para. 19.

70 Ibid. para. 18.

71 [2008] EWHC 2319 (Comm).

72 Case C-45/96 [1998] ECR I-1199.

73 [2008] EWHC 2319 (Comm), para. 28.

74 QBD, unreported, 17 June 2005.

operating closer to the actual wording of the Regulations, held that the Regulations did not apply to surety transactions and his view has subsequently been described as "compelling"[75] and "convincing".[76]

This unevenness of approach is also evidenced by the Judgment of the Supreme Court in *OFT* v *Abbey National plc*. Although the Supreme Court, unsurprisingly, accepted that a purposive approach to interpretation should be adopted in relation to the Regulations,[77] the extent to which it successfully did so is debatable. For example, Lord Walker (with whom Lord Phillips, Lady Hale and Lord Neuberger expressed agreement), whilst (entirely appropriately) referring to both the Regulations and the text of the Directive, seems to have adopted a largely literal interpretation of the relevant provisions. Thus, His Lordship stated:

> When one turns to the other part of the quid pro quo of a consumer contract, the price or remuneration, the difficulty of deciding which prices are essential is just the same, and Regulation 6(2)(b) contains no indication that only an "essential" price or remuneration is relevant. Any monetary price or remuneration payable under the contract would naturally fall within the language of paragraph (b) (I discount the absence of a reference to part of the price or remuneration for reasons already mentioned).[78]

It is, of course, true that Lord Walker later noted that he believed his conclusions were in line with the *travaux préparatoires*.[79] In so doing, Lord Walker made a distinction between consumer protection and consumer choice; and referred, with apparent approval,[80] to an article by Professor Hugh Collins[81] which stated:

> The Directive does not require consumer contracts to be substantively fair, but it does require them to be clear. Clarity is essential for effective market competition between terms. What matters primarily for EC contract law is consumer choice, not consumer rights.[82]

Yet, with respect (and ignoring Lord Walker's apparent, possibly uneasy, conflation of "consumer choice" and "clarity"), such a view does not seem to do full justice to the different facets of the Directive (and consequently the Regulations). More specifically, the reach of the Directive and the Regulations is not limited to terms which are not in plain intelligible language;[83] nor is it possible to say that substantive unfairness is not relevant under the Directive or the Regulations (although the *extent* to which it is relevant might be, of course, intensely debated[84]).[85] The backdrop to this discussion is, of course, the much-debated[86] issue of the purpose, and, indeed, the extent of the competence, of EU legislation in the consumer arena. Over the years various bases for EU legislation in this area

75 *Manches LLP* v *Carl Freer* [2006] EWHC 991, para. 25, per Judge Philip Price QC.

76 *Williamson* v *Governor of the Bank of Scotland* [2006] EWHC 1289, para. 46, per George Bompas QC, sitting as a deputy judge.

77 *OFT* v *Abbey National plc* [2009] UKSC 6, see, for example, para. 38, per Lord Walker.

78 Ibid. para. 41.

79 Ibid. para. 44.

80 Ibid.

81 H Collins, "Good faith in European contract law" (1994) 14 *OJLS* 229.

82 Ibid. p. 238.

83 See Regulation 5.

84 Nor is the interplay between substantive and procedural unconscionability under the Regulations not unproblematic following *Director General of Fair Trading* v *First National Bank plc* [2001] UKHL 52 above.

85 See, for example, Unfair Terms in Consumer Contracts Regulations 1999, Sch. 2.

86 Cf. M Kenny "The 2004 Communication on European Contract Law: those magnificent men in their unifying machines" (2005) 30 *European Law Review* 724.

have been advanced, such as establishing an internal market or preventing the distortion of competition.[87] Yet none of those bases proved entirely unproblematic[88] and, more recently, the desire to promote consumer confidence in the internal market has begun to emerge as a key base of EU intervention in the area of consumer law.[89] This is, for example, evidenced by the prominent appeal to "consumer confidence" notions in the recent proposed Consumer Rights Directive:[90]

> These disparities create significant internal market barriers affecting business and consumers. They increase compliance costs to business wishing to engage in cross border sale of goods or provision of services. Fragmentation also undermines consumer confidence in the internal market. The negative effect on consumer confidence is strengthened by an uneven level of consumer protection across the Community. This problem is particularly acute in the light of new market developments.[91]

The key point for present purposes is that it must, at least, be arguable that "consumer confidence" (even when counter-balanced by business interests) is not fostered solely through making sure contractual terms are in "plain intelligible language".[92]

Lord Mance's judgment (with which Lady Hale and Lord Neuberger expressed agreement) traces the gestation of the Directive;[93] and Lord Mance places heavy reliance[94] on an influential article by Professors Brandner and Ulmer[95] to conclude:[96]

> In my opinion, the identification of the price or remuneration for the purposes of . . . regulation 6(2) is a matter of objective interpretation for the court. The court should no doubt read and interpret the contract in the usual manner, that is having regard to the view which the hypothetical reasonable person would take of its nature and terms. But there is no basis for requiring it to do so by attempting to identify a "typical consumer" or by confining the focus to matters on which it might conjecture that he or she would be likely to focus. The consumer's protection under the Directive and Regulations is the requirement of transparency on which both insist. That being present, the consumer is to be assumed to be capable of reading the relevant terms and identifying whatever is objectively the price and remuneration under the contract . . .

87 C Twigg-Flesner, *The Europeanisation of Contract Law* (London: Routledge-Cavendish 2008), pp. 25ff.

88 Ibid.

89 Cf. T Wilhelmsson, "The abuse of the 'confident consumer' as a justification for EC consumer law" (2004) 27 *Journal of Consumer Policy* 317.

90 See, generally, G Howells and R Schulze, *Modernising and Harmonising Consumer Contract Law: With Reference to the Planned Horizontal Consumer Contract Directive* (Munich: Sellier 2009). The original proposed Directive is available at: http://ec.europa.eu/consumers/rights/docs/COMM_PDF_COM_2008_0614_F_EN_PROPOSITION_DE_DIRECTIVE.pdf.

91 Recital (7) of the original proposal. See also *EU Consumer Policy Strategy 2007–2013: Empowering consumers, enhancing their welfare, effectively protecting them*, COM (2007) 99 Final.

92 Cf., for example, Recital (13) and Recital (16) of Directive 2005/29/EC concerning unfair business-to-consumer commercial practices in the internal market.

93 *OFT* v *Abbey National plc* [2009] UKSC 6, paras 109–112.

94 Ibid. para. 109.

95 H E Brandner and P Ulmer, "The Community Directive on Unfair Terms in Consumer Contracts: some critical remarks on the proposal submitted by the EC Commission" (1991) 28 *CMLR* 647.

96 *OFT* v *Abbey National plc* [2009] UKSC 6, para. 113.

Yet Lord Mance does not really unpack the coordinates within which this hypothetical, and reasonable, person is to operate;[97] how is this person to filter price from non-price terms? Moreover, in so far as Lord Mance's conclusions are based on the view that the protection afforded to consumers under the Directive and the Regulations is one of transparency, it is, with respect (and not necessarily meaning to suggest that "transparency" and "clarity" are coterminous), open to some of the same criticisms as Lord Walker's conclusion on this point. In particular, it is not possible to state that the unfairness test under the Directive and the Regulations is limited to questions of transparency.[98] Moreover, again, it must, at least, be arguable that "consumer confidence" objectives (even when counter-balanced by business interests) are not fostered solely through making sure contractual terms are "transparent".[99]

Notwithstanding the foregoing criticisms, it must be acknowledged that the task of interpreting, in an appropriate fashion, Regulation 6(2) is not a straightforward task. In particular, the Supreme Court was faced with an ambiguous "compromise" provision in a directive; the purpose(s) of the Directive was debatable; the European Court of Justice had essentially not considered the relevant provision; and the provision has been transposed in different ways throughout the EU.[100] In such circumstances, a robust and efficient reference process to the European Court of Justice, to act as a compass, might have been helpful; yet, it seems that the reference process is not always perceived to have such qualities[101] (even if the Supreme Court had been inclined to make a reference to the European Court of Justice).[102]

The resultant interpretation of Regulation 6(2)(b) by the Supreme Court

As noted above, the Supreme Court was unanimous in finding that the charges in question were covered by Regulation 6(2)(b); thus precluding, at least, any challenge to the terms based merely on the assertion that these charges were excessive.[103] Nevertheless, it is possible to identify some differences of construction of Regulation 6(2)(b) within the Supreme Court. For example, Lord Walker, in effect, construed Regulation 6(2)(b) as referring to *some* forms of *monetary* consideration payable under the relevant contract.[104] Such a construction is noteworthy for, at least, two reasons. First, even when heeding Lord Steyn's note of caution in *Director General of Fair Trading* v *First National Bank plc*,[105] it implies that it is possible to review some central *non-monetary* payment terms for substantive fairness. Thus, for example, if a consumer part-exchanges their old home for a new home, on Lord Walker's view of Regulation 6(2)(b) it would seem that standard terms relating to how the "credit" for the old home is to be determined might be subject to review, under the Regulations, for substantive fairness. Yet *if*, as Lord Walker believes,[106] the Regulations are

97 Compare, in general terms, A Chandler and J Devenney, "Mistake as to identity and the threads of objectivity" (2004) *Journal of Obligations and Remedies* 7.

98 See, for example, Sch. 2, 1(l).

99 Again compare, for example, Recital (13) and Recital (16) of Directive 2005/29/EC concerning unfair business-to-consumer commercial practices in the internal market.

100 H Schulte-Nölke, *EC Consumer Law Compendium: Comparative Analysis* (2008), p. 345, available at: http://ec.europa.eu/consumers/rights/docs/consumer_law_compendium_comparative_analysis_en_final.pdf.

101 Cf. C Turner and R Munoz, "Revisiting the judicial architecture of the European Union" (1999/2000) 19 *YBEL* 1.

102 On which see the text to n. 169 below.

103 See, for example, *OFT* v *Abbey National plc* [2009] UKSC 6, para. 51, per Lord Walker.

104 Ibid. para. 43.

105 [2001] UKHL 52, para. 34.

106 *OFT* v *Abbey National plc* [2009] UKSC 6, para. 44.

concerned with providing consumers with clarity rather than protection per se, it is difficult to maintain his distinction between monetary and non-monetary standard terms.

The second aspect of Lord Walker's construction of Regulation 6(2)(b) which is noteworthy is his view that only *some* forms of *monetary* consideration payable under the relevant contract are covered by that provision. Of course, even if Lord Walker had been considering this provision afresh, it would have been difficult for him not to have made this concession given some of the clauses listed in Schedule 2 as indicative of unfair terms.[107] Yet Regulation 6(2)(b) had already, in effect,[108] been considered by, inter alia, the House of Lords in *Director General of Fair Trading* v *First National Bank plc*.[109] The key issue in that case for present purposes was whether the predecessor to Regulation 6(2)(b) covered a term in a credit agreement which provided that interest, at the contractual rate, was to continue to accrue "after as well as before any judgment (such obligation to be independent of and not to merge with the judgment)". The House of Lords unanimously held that this term did not come within the predecessor of Regulation 6(2)(b). In so doing, Lord Steyn noted:

> Clause 8 of the contract, the only provision in dispute, is a default provision. It prescribes remedies which only become available to the lender upon the default of the consumer. For this reason the escape route of regulation 3(2) is not available to the bank. So far as the description of terms covered by regulation 3(2) as core terms is helpful at all, I would say that clause 8 of the contract is a subsidiary term. In any event, regulation 3(2) must be given a restrictive interpretation. Unless that is done regulation 3(2)(a) will enable the main purpose of the scheme to be frustrated by endless formalistic arguments as to whether a provision is a definitional or an exclusionary provision. Similarly, regulation 3(2)(b) dealing with "the adequacy of the price or remuneration" must be given a restrictive interpretation. After all, in a broad sense all terms of the contract are in some way related to the price or remuneration. That is not what is intended . . . It would be a gaping hole in the system if such clauses were not subject to the fairness requirement. For these further reasons I would reject the argument of the bank that regulation 3(2), and in particular 3(2)(b), take clause 8 outside the scope of the Regulations.[110]

In *Abbey*, Lord Walker used this passage to qualify his view that Regulation 6(2)(b) covers *monetary* consideration payable under the relevant contract; Lord Walker was apparently of the view that default provisions were not captured by Regulation 6(2)(b). Leaving aside the conundrum of why, for example, excessive default terms are covered by the Regulations whereas excessive non-default terms are not (and the fact that *Director General of Fair Trading* v *First National Bank plc* seemed to involve a *primary* obligation which was extended into the default zone by clause 8), this construction is intriguing. In particular, it is not difficult to conceive of situations where an enterprise with many customers relies on the statistical probability of default by individual customers (and the associated default charges) as part of its profitability calculations and forecasts.[111] Indeed, it may be that bank charges were at one time regarded as default clauses; whereas today, certainly in the aftermath of *Abbey*, they can be regarded as non-default clauses given, for example, the reliance which banks place upon them (and possibly different attitudes towards credit and debt).[112] Thus, Lord Phillips noted:

107 See, for example, Sch. 2, 1(e).

108 Under the then Unfair Terms in Consumer Contracts Regulations 1994 (1994/3159).

109 [2001] UKHL 52.

110 Ibid. para. 34.

111 Cf. in broad terms *UK Housing Alliance (North West) Ltd* v *Francis* [2010] EWCA Civ 117.

112 On which see, generally, C Scott and J Black, *Cranston's Consumers and the Law* (London: Butterworths 2000), pp. 229–32.

> When the relevant facts are viewed as a whole, it seems clear that the Relevant Charges are not concealed default charges designed to discourage customers from overdrawing on their accounts without prior arrangement. Whatever may have been the position in the past, the Banks now rely on the Relevant Charges as an important part of the revenue that they generate from the current account services. If they did not receive the Relevant Charges they would not be able profitably to provide current account services to their customers in credit without making a charge to augment the value of the use of their funds.[113]

Returning to Lord Walker's judgment, it is also clear that His Lordship did not qualify his construction of Regulation 6(2)(b) *solely* by reference to default terms. Indeed, such a view would have been clearly unsustainable given that some of the terms listed in Schedule 2 (as indicative unfair terms) are monetary obligations which are not necessarily default terms.[114] Again *if*, as Lord Walker believes,[115] the Regulations are concerned with providing consumers with clarity rather than protection per se, it is difficult to maintain his distinction between different types of monetary standard terms.

Lord Walker provides little guidance on which monetary obligations come within Regulation 6(2)(b) and which monetary obligations do not do so; indeed, somewhat ironically after agreeing with a description of the Court of Appeal's approach as "over-elaborate",[116] Lord Walker reverts to language (admittedly with a "health-warning")[117] such as "ancillary" and "non-core".[118] One might argue that this uncertainty may undermine "consumer confidence"; yet, it must be debatable the extent to which, if at all, the minutiae of consumer protection in a particular Member State (as opposed to the broad lines of consumer protection in any given Member State) actually affects "consumer confidence".[119] On the other hand, such uncertainty might impact on business patterns,[120] and may also inefficiently expend some of the finite resources that bodies such as the OFT have for work in this area.

In broad agreement with Lord Walker, Lord Mance was of the opinion that not all of the payment provisions under a relevant contract are necessarily covered by Regulation 6(2)(b);[121] and he, like Lord Walker, gave default provisions as an example of payment provisions which, seemingly, are not captured by Regulation 6(2)(b).[122] As such, this aspect of Lord Mance's position attracts some of the same criticisms as Lord Walker's position. Yet, as noted above, Lord Mance's ultimate conclusion on Regulation 6(2)(b) has a rather different flavour to Lord Walker's conclusions; in particular, Lord Mance regards the determination of the "price or remuneration" as the province of the (coordinates lacking) "hypothetical reasonable person". The difficulty with such an approach, of course, is that – as is vividly illustrated by this litigation – the viewpoint of the banks and the viewpoint of their consumers on this issue is likely to be markedly different; and there may be little ground for the "hypothetical reasonable person" to steer a middle course. Ultimately,

113 *OFT* v *Abbey National plc* [2009] UKSC 6, para. 88.

114 See, for example, Sch. 2, 1(d).

115 *OFT* v *Abbey National plc* [2009] UKSC 6, para. 44.

116 Ibid. para. 38.

117 Ibid. para. 46.

118 Ibid.

119 Cf. R Goode, "Contract and commercial law: the logic and limits of harmonisation" in F W Grosheide and E Hondius, *International Contract Law* (Antwerp: Intersentia 2004).

120 See, for example, Recital (8) of the proposed Consumer Rights Directive.

121 *OFT* v *Abbey National plc* [2009] UKSC 6, para. 113.

122 Ibid.

therefore, Lord Mance's construction of Regulation 6(2)(b) may also result in uncertainty which impacts on business patterns,[123] and possibly inefficiently expends some of the finite resources that bodies such as the OFT have for work in this area.

The interpretation of regulation 6(2)(a) by the Supreme Court

Although the attention of the Supreme Court in *Abbey* was focused on Regulation 6(2)(b), the Supreme Court did make some interesting comments on the proper interpretation of Regulation 6(2)(a). As readers will, no doubt, remember, Regulation 6(2) provides "[i]n so far as it is in plain intelligible language, the assessment of fairness of a term shall not relate . . . (a) to the definition of the main subject matter of the contract"; and some indication of the intended operation of Regulation 6(2)(a) can be found in Recital (19):

> Whereas, for the purposes of this Directive, assessment of unfair character shall not be made of terms which describe the main subject matter of the contract nor the quality/price ratio of the goods or services supplied; whereas the main subject matter of the contract and the price/quality ratio may nevertheless be taken into account in assessing the fairness of other terms; whereas it follows, inter alia, that in insurance contracts, the terms which clearly define or circumscribe the insured risk and the insurer's liability shall not be subject to such assessment since these restrictions are taken into account in calculating the premium paid by the consumer.

One of the immediate difficulties that arises in relation to Regulation 6(2)(a) is that there is, prima facie, an *argument* that it renders exclusion clauses immune for review under the Regulations. More specifically, there is an argument that exclusion clauses define the obligations, and risks, of the parties (rather than merely exclude liability)[124] and, if so, it follows from Recital (19) that they are caught by Regulation 6(2)(a). Yet, whatever the true nature of exclusion clauses, it is clear that exclusion clauses are often regarded as prime candidates for the label of "unfair terms" under the Regulations.[125]

Nevertheless difficulties do still persist. Recital (19) makes it clear that "terms which clearly define or circumscribe the insured risk and the insurer's liability" are captured by Regulation 6(2)(a); and it also seems beyond argument that exclusion clauses are not caught by Regulation 6(2)(a); but what about other types of exclusion? For example, in surety transactions the surety has the right to be discharged on the occurrence of certain events; thus the creditor has a well-established equitable duty not to release any security which it holds in relation to the principal debt.[126] As a result it is commonplace[127] for creditors to insert clauses into surety transactions which exclude a surety's right to be discharged on the happening of such events.[128] The key point to note about such clauses is that they aim to preserve the surety's liability rather than to exclude the creditor's liability.[129] Nevertheless, it is difficult to predict whether or not such clauses would fall within Regulation 6(2)(a).[130]

123 Again see, for example, Recital (8) of the proposed Consumer Rights Directive.

124 See, for example, the speech of Lord Diplock in *Photo Production Ltd* v *Securicor Transport Ltd* [1980] AC 827. Cf. D Yates, *Exclusion Clauses in Contracts* 2nd edn (London: Sweet & Maxwell 1992), pp. 123–33.

125 See, generally, J Beatson, A Burrows and J Cartwright, *Anson's Law of Contract* (Oxford: OUP 2010), p. 207.

126 See, for example, *Skipton Building Society Ltd* v *Stott* [2001] QB 261.

127 See *Barclays Bank plc* v *Kufner* [2008] EWHC 2319 (Comm), para. 16.

128 See further J Devenney, "Aspects of property, security and guarantees" in M Furmston and J Chuah, *Commercial and Consumer Law* (London: Pearson 2010), p. 42.

129 See McCormack, "Protection of surety guarantors", n. 61 above, p. 171.

130 The point does not seem to have been raised in the relevant case law.

In *Abbey*, Lord Walker read Regulation 6(2)(a) and Regulation 6(2)(b) as representing the essential *quid pro quo*[131] present in the relevant consumer contracts although he stopped short of stating that Regulation 6(2) should be read either conjunctively or disjunctively.[132] Moreover, Lord Walker (seemingly in contradistinction to Lord Mance)[133] interpreted the phrase "main subject matter" of the contract in a rather generic fashion:

> The main subject-matter may be goods or services. If it is goods, it may be a single item (a car or a dishwasher) or a multiplicity of items. If for instance a consumer orders a variety of goods from a mail-order catalogue . . . there is no possible basis on which the court can decide that some items are more essential to the contract than others. The main subject matter is simply consumer goods ordered from a catalogue . . . Similarly, a supply of services may be simple (an entertainer booked to perform for an hour at a children's party) or composite (a week's stay at a five-star hotel offering a wide variety of services). Again, there is no principled basis on which the court could decide that some services are more essential to the contract than others and again the main subject matter must be described in general terms—hotel services.[134]

A number of points may be made in relation to this statement. The first point to make is that, initially at least, it seems that Lord Walker's "generic" approach to Regulation 6(2)(a) – with its high level of abstraction – is simple to apply. Yet, on reflection, it is not clear that this is necessarily so. Again we may use the thorny question of whether or not the Regulations apply to non-professional surety transactions as an example.[135] As is well known, doctrinally (and in general terms) a surety assumes secondary and accessory liability.[136] In other words, the surety's obligations are generally dependent on the existence of a primary obligation between the creditor and primary debtor,[137] and on the default of the primary debtor.[138] As is also well known, under the general law there are also various well-established situations where a surety's obligations will be discharged (we have already referred to the creditor's well-established equitable duty not to release any security which it holds in relation to the principal debt[139]).[140] Nevertheless, as we have seen,[141] creditors often seek to "exclude" the surety being discharged in such circumstances. However, such clauses – and indeed "principal debtor" clauses[142] – do not necessarily lead to the conclusion that the contract is not a contract of suretyship.[143] Yet, on Lord Walker's generic approach to Regulation 6(2)(a), how would the main subject matter of such contracts be

131 Although cf. His Lordship's statement (*OFT* v *Abbey National plc* [2009] UKSC 6, para. 46): ". . . delivery of goods or peripheral extras may be disregarded as ancillary for the purposes of para (a) of Regulation 6(2), but the charges for them, if payable under the same contract, are part of the price for the purposes of para (b)."

132 Ibid. para. 31.

133 Ibid. para. 103, Lord Mance noted: ". . . since the deposit with or transfer to a bank of money is the main or part of the main subject matter of a banking contract, any assessment of the fairness of it or its legal consequences would appear to be excluded under regulation 6(2)(a), rather than (b)".

134 Ibid; paras 39–40.

135 See McCormack, "Protection of surety guarantors", n. 61 above, pp. 172–3.

136 See further Devenney, "Aspects of property", n. 128 above, p. 24.

137 See L S Sealy and R J A Hooley, *Commercial Law* 4th ed (Oxford: OUP 2008), p. 1150.

138 See *Moschi* v *Lep Air Services Ltd* [1973] AC 331.

139 See, for example, *Skipton Building Society Ltd* v *Stott* [2001] QB 261.

140 See further Devenney, "Aspects of property", n. 128 above, pp. 46–8.

141 See *Barclays Bank plc* v *Kufner* [2008] EWHC 2319 (Comm), para. 16.

142 See, for example, *Heald* v *O'Connor* [1971] 1 WLR 497 and *General Produce Co* v *United Bank Ltd* [1979] 2 Lloyd's Rep 255.

143 See further Devenney, "Aspects of property", n. 128 above, p. 27.

framed? More specifically would the main subject matter of such contracts be the provision of surety *stricto sensu*, or (at a higher level of abstraction) merely the provision of security? If the former approach is to be preferred over the latter approach then (subject to what follows) it is, perhaps, more likely that the Regulations do apply to non-professional surety transactions; and this question is not just one of academic indulgence. Such transactions perform an important economic role[144] and the application of the Regulations to such transactions may significantly alter the balance of interests between the surety and the creditor,[145] possibly resulting in suretyship transactions becoming less attractive to banks[146] and thus causing a narrowing in access to credit.[147]

The second point to make (in respect of Lord Walker's "generic" approach to Regulation 6(2)(a)) is that it is clear – from the epithet "main" – that not all terms which describe the subject matter of the contract come within Regulation 6(2)(a). A question, therefore, arises as to how, in this context, the *main* subject matter of a contract is distinguished from the *remainder* of the subject matter of the contract; and this is a question upon which Lord Walker's judgment is less than helpful. Indeed, Lord Walker, again, refers to those terms not captured by Regulation 6(2) as "ancillary" terms,[148] and, as such, is subject to some of the same criticisms outlined above in relation to Regulation 6(2)(a).[149]

The third point relates to the issue of whether or not Lord Walker's rather "generic" approach to part of Regulation 6(2)(a) should prevail throughout that (sub-) provision. More specifically, which terms form part of "*the definition* of the main subject"[150] for the purpose of Regulation 6(2)(a)? Readers will no doubt be familiar with the somewhat

144 For example, in *Royal Bank of Scotland* v *Etridge (No 2)* [2001] UKHL 44, paras 34–35, Lord Nicholls stated: "The problem . . . raised by the present appeals is of comparatively recent origin. It arises out of the substantial growth in home ownership over the last 30 or 40 years . . . More than two-thirds of householders in the United Kingdom now own their own homes. For most home-owning couples, their homes are their most valuable asset. They must surely be free, if they so wish, to use this asset as a means of raising money, whether for the purpose of the husband's business or for any other purpose . . . Bank finance is in fact by far the most important source of external capital for small businesses with fewer than ten employees. These businesses comprise about 95 percent of all businesses in the country, responsible for nearly one-third of all employment . . . If the freedom of home-owners to make economic use of their homes is not to be frustrated, a bank must be able to have confidence that a wife's signature of the necessary guarantee and charge will be as binding upon her as is the signature of anyone else on documents which he or she may sign. Otherwise banks will not be willing to lend money on the security of a jointly owned house or flat." See, generally, J Devenney and M Kenny, "Unfair terms, surety transactions and European harmonisation: a crucible of Europeanised private law?" (2009) *Conveyancer and Property Lawyer* 295.

145 G Andrews and R Millett, *The Law of Guarantees* 4th edn (London: Thompson 2005), argue (at p. 85) that: "If the regulations are applied to bank guarantees, it will be seen that there is considerable scope for an interventionist judiciary to redress the balance between creditor and surety significantly." Of course, the orthodox view is that before the Regulations the Law of England and Wales was unconcerned with the substantive fairness of surety transactions. Yet this is, perhaps, to overstate the case. As I have argued elsewhere, the doctrine of undue influence is the mainstay of the protection afforded to sureties: see J Devenney, L Fox-O'Mahony and M Kenny, "Standing surety in England and Wales: the Sphinx of procedural protection" (2008) *Lloyds Maritime and Commercial Law Quarterly* 527. In cases of undue influence, it is sometimes tempting to view the former manifest disadvantage requirement solely in evidential terms. Nevertheless, if undue influence is based on a notion of unconscionability (on which see J Devenney and A Chandler, "Unconscionability and the taxonomy of undue influence" (2007) *Journal of Business Law* 541) the relevance of the substantive fairness of the transaction will not merely be evidential: see, for example, *Dunbar Bank plc* v *Nadeem* [1998] 3 All ER 876.

146 See M Kenny, "Standing Surety", n. 66 above, pp. 195–6.

147 See, generally, J Devenney and M Kenny, "Unfair terms", n. 144 above.

148 *OFT* v *Abbey National plc* [2009] UKSC 6, para. 46.

149 See J Devenney and M Kenny, "Unfair terms", n. 144 above.

150 Emphasis added.

"generic" (and, indeed, controversial)[151] approach adopted by the House of Lords in *Ashington Piggeries Ltd* v *Christopher Hill Ltd*[152] to the question of which descriptive words form part of the description of goods for the purposes of s. 13 of the Sale of Goods Act 1979. If a similar approach were adopted in this context, the scope of the exception in Regulation 6(2)(a) would be much curtailed and, potentially, consumer protection (and, possibly, consumer confidence) would be increased; for example, if a consumer purchases a painting believing it to be by a famous painter (and pays a price which reflects this fact), a term in the contract which, in effect (and assuming it is not an exclusion clause), states that the seller does not warrant that the painting is by the famous painter in question might not be caught by the exception in Regulation 6(2)(a) if an *Ashington Piggeries*-like approach is adopted.

Overall our analysis of Regulation 6(2)(a), in the light of the comments of the Supreme Court in *Abbey* on that provision, serves to highlight the uneasy compromise made in Article 4(2) of the Directive against the backdrop of a not entirely coherent legislative basis. Moreover, our discussion also underlines the fact that a (largely) passively reproductive[153] approach to transposing EU legislation places the onus of untangling any knots, and maintaining some coherence in domestic law, firmly on the doorstep of the judiciary.

ECJ reference?

As noted above, the Supreme Court in *Abbey* was faced with an ambiguous "compromise" provision in a directive; the purpose(s) of the Directive was debatable; the European Court of Justice had essentially not considered the relevant provision; and the provision has been transposed in different ways throughout the EU.[154] In such circumstances it must, at least, be arguable that an interpretative reference to the European Court of Justice was required in *Abbey*.[155] Yet the Supreme Court, with variable assurance (and some controversy),[156] felt that it was unnecessary to make such a reference. Spearheading this stance was the view of Lord Walker[157] (which was essentially supported by Lord Mance[158] and Lady Hale)[159] that the issue was *acte clair*. However, as even Lord Walker came close to conceding,[160] such a view is difficult to sustain given the differing views of "very experienced"[161] judges in the course of this litigation; and, one might add, the views of judges in other cases (and, indeed, in *Abbey* Lord Phillips (supported to some extent by Lord Neuberger)[162] did not find the issue *acte clair*).[163] Perhaps as a result of these difficulties, Lord Mance sought to buttress the *acte clair* argument by stating:

151 See, for example, I Brown, *Commercial Law* (London: Butterworths 2000), pp. 457–8.

152 [1972] AC 441.

153 See n. 15 above.

154 Schulte-Nölke, *EC Consumer Law Compendium*, n. 100 above.

155 Cf. P Davies, "Bank charges in the Supreme Court' (2010) 69 *Cambridge Law Journal* 21, p. 23.

156 Ibid.

157 *OFT* v *Abbey National plc* [2009] UKSC 6, paras 48–50.

158 Ibid. paras 115–17.

159 Ibid. para. 92.

160 Ibid. para. 49.

161 Ibid.

162 Ibid. para. 120.

163 Ibid. para. 91.

> In the present case, we are concerned with a relatively simple sentence, using simple and basic concepts, and the scope for different readings of different language texts seems very limited.[164]

The task, of course, may be relatively straightforward if a literal-leaning approach is to be adopted; yet, as noted above, Article 4(2) of the Directive is an uneasy compromise provision, the purpose(s) for which is debatable; and, given the fact that internal market goals may evolve as the internal market itself evolves,[165] the fact that a provision has been transposed in different ways throughout the EU is not without significance.

A secondary argument, advanced by Lords Walker and Mance for not referring the matter to the European Court of Justice, relied on a distinction between the *construction* of Article 4(2) and the *application* of Article 4(2); they argued, with some justification,[166] that the former was a matter for EU law whereas the latter was a matter for domestic law.[167] More specifically, they argued that *even if* the Court of Appeal were correct in its construction of Article 4(2), the Court of Appeal was incorrect in its application of this construction of Article 4(2) to the facts;[168] accordingly, so the argument went, the correct interpretation of Article 4(2) was not central to the determination of the appeal and, therefore, a reference to the European Court of Justice was not necessary.

Such reasoning is, of course, superficially attractive. Yet, it is not clear that this reasoning entirely separates questions of law from questions of fact; more specifically the conclusion that the Court of Appeal wrongly applied its interpretation of Article 4(2) is dependent on *a particular view* of the core/ancillary-terms dichotomy which, surely, is partly a question of law. There is also a wider issue: this argument effectively concedes that the Supreme Court may have been wrong on the question of interpretation!

Whatever the "formal" reasons for concluding that a reference to the European Court of Justice was not necessary, it seems that there were wider considerations at play. In particular, the Supreme Court seemed concerned by the delay which a reference might entail.[169] This was, of course, not the first time that concerns about the efficiency of the reference process had been raised. For example, in *Page* v *Combined Shipping and Trading Co. Ltd*[170] Staughton LJ stated:

> We have been shown the French, German and Italian versions all of which use the word "normal/normale" instead of "proper". That does not necessarily mean the same as "normal" in English; similarities in language can be deceptive. It seems to me, in short, that we ought to conclude that Mr Page has a good arguable case for recovering a substantial sum. It may well be that when this comes to trial we shall have to refer the problem to the European court, and it will take another two years after that before a decision emerges as to what the

164 *OFT* v *Abbey National plc* [2009] UKSC 6, para. 115.

165 See also the comment of Bingham J in *Customs and Excise Commissioners* v *Samex* [1983] 3 CMLR 194, para. 31, that the ECJ "has a panoramic view of the Community and its institutions, a detailed knowledge of the treaties and of much subordinate legislation made under them, and an intimate familiarity with the functioning of the Community market which no national judge denied the collective experience of the Court of Justice could hope to achieve".

166 Cf. *Freiburger Kommunalbauten GmbH Baugesellschaft & Co KG* v *Ludger Hofstetter and Ulrike Hofstetter* [2004] ECR-I 3403, para. 22, where the European Court of Justice noted that it "may interpret general criteria used by the Community legislation in order to define the concept of unfair terms. However, it should not rule on the application of these general criteria to a particular term."

167 *OFT* v *Abbey National plc* [2009] UKSC 6, paras 50 and 116 respectively.

168 Ibid.

169 See, for example, ibid. para. 50, per Lord Walker (with whom Lady Hale and Lord Neuberger agreed).

170 [1996] CLC 1952.

> regulation really means. Maybe the parties will think there are better methods of spending their time and their money than disputing that for a long period of time. But for the present it is enough for us to say, I think, that there is a good arguable case.[171]

Yet, *even if* such considerations are legitimate as a matter of EU Law,[172] it is somewhat ironic that the Supreme Court argued – as a reason not to make a reference to the European Court of Justice – that there was a "public interest" in resolving the issue quickly given, for example, the number of stayed cases awaiting its decision:[173] we have already seen, that by focusing on a very narrow issue whilst opaquely hinting[174] that there may be a possibility to challenge the charges in question under the Regulations, the Supreme Court's decision is less than helpful from a case management point of view! But were there other drivers behind the Supreme Court's decision?

The impact of the banking crisis?

The Supreme Court ruled on the *Abbey* appeal in the immediate aftermath of the banking crisis; and one does not have to be a conspiracy-theorist to wonder whether or not a (perhaps perceived) vulnerability on the part of banks[175] (and the associated impact on the economy)[176] in any way impacted (not necessarily illegitimately) on the Supreme Court's ruling in *Abbey*.[177] Of course, we will probably never know for sure but the consequences of a ruling against the banks in *Abbey* do make for stark reading: Lord Walker noted that in 2006 alone the banks (by which he presumably meant the banks involved in the *Abbey* appeal) made approximately £2.56 billion from these charges[178] (which equated to approximately one-third of the revenue made on current accounts);[179] "many thousands"[180] of cases against banks on this issue were stayed in the County Courts pending a decision in *Abbey*; and, presumably, the limitation clock would only begin ticking once the Supreme Court handed down its judgment.[181]

---

171 [1996] CLC 1952, p. 1956.

172 Cf. Twigg-Flesner, *The Europeanisation of Contract Law*, n. 87 above, pp. 130–3.

173 Lord Walker suggested that there may be thousands of such cases: *OFT* v *Abbey National plc* [2009] UKSC 6, para. 17.

174 See, for example, ibid. para. 91, per Lord Phillips.

175 See M Kenny, J Devenney and L. Fox-O'Mahony, "Conceptualising unconscionability in Europe: in the kaleidoscope of private and public law" in M Kenny, J Devenney and L Fox-O'Mahony, *Unconscionability in European Private Financial Transactions: Protecting the Vulnerable* (Cambridge: CUP 2010), p. 377, at p. 378.

176 Cf. Wishart, "Transparency and fairness", n. 27 above, p. 158.

177 Ibid.

178 *OFT* v *Abbey National plc* [2009] UKSC 6, para. 36.

179 Ibid. para. 47.

180 Ibid. para. 17, per Lord Walker.

181 See Limitation Act 1980, s. 32(1) and *Kleinwort Benson Ltd* v *Lincoln CC* [1999] 2 AC 349. Indeed, one is reminded of the words of Lord Browne-Wilkinson in that case, at p. 363: "Much commercial and property activity occurs on the basis of law which is not laid down by judicial decision. Such 'law' consists of the practice and understanding of lawyers skilled in the field. If, before payment, the payer had sought advice in some cases he would have been told that the law was dubious: if having received such advice he paid over, he must have taken the risk that the law was otherwise and cannot subsequently recover what he has paid. In other cases, he would have been told that the law was clear and he could safely act on it. If in this latter case the payer acted on the law as so advised and subsequently a court held that the law was not as advised, can the payer recover his payment as moneys paid under a mistake of law? In the ordinary case, the payer's adviser will just have given wrong legal advice: as a result the payment will have been paid under a mistake of law and will be recoverable. But in a limited number of cases, of which this may be one, it is not really possible to say that the legal adviser made a mistake in advising as he did . . . It used to be said that the practice of conveyancers of repute was strong evidence of real property law: see *In re Hollis' Hospital Trustees and Hague's*

Implications for EU harmonisation

As suggested above, there are various lessons in *Abbey* for any EU harmonisation agenda: lessons relating the role of the judiciary in EU harmonisation;[182] lessons for the EU harmonisation agenda on the role of the legislatures in individual Member States; the need for an efficient and robust reference procedure to the European Court of Justice, and so forth. There is also a wider issue relating to the exercise of determining whether or not a particular clause is "unfair". As noted above, the key provision is Regulation 5:

> A contractual term which has not been individually negotiated shall be regarded as unfair if, contrary to the requirement of good faith, it causes a significant imbalance in the parties' rights and obligations arising under the contract, to the detriment of the consumer.

An insight into (part of) this exercise can be found in *Director General of Fair Trading* v *First National Bank plc*[183] where Lord Bingham stated:

> Good faith in this context is not an artificial or technical concept; nor, since Lord Mansfield was its champion, is it a concept wholly unfamiliar to British lawyers. *It looks to good standards of commercial morality and practice.*[184]

The important point for present purposes is that this exercise appears to be loaded with social, economic and behavioural norms and, as I have argued elsewhere,[185] there are clear differences in such norms throughout the EU in the broad context of financial transactions. For example, one might refer to the significant differences in the operation of the doctrine of undue influence, in the context of non-professional surety protection, when comparing Ireland with England and Wales.[186]

Returning to *Abbey*, the Supreme Court did not state that the charges in question could never be challenged under the Regulations. As Lord Phillips stated:

> I do not believe any challenge to the fairness of the Relevant Terms has been made on the basis that they cause the overall package of remuneration paid by those in debit to be excessive having regard to the package of services received in exchange. In these circumstances the basis on which I have answered the narrow issue would seem to render that issue academic. It may be that, if and when the OFT challenges the fairness of the Relevant Terms, issues will be raised that ought to be referred to Luxembourg. That stage has not yet been reached.[187]

---

181 [cont.] *Contract* [1899] 2 Ch. 540, 551 . . . [and] Denning L.J. said in *In re Downshire Settled Estates; Marquess of Downshire v. Royal Bank of Scotland* [1953] Ch. 218 , 279: 'The practice of the profession in these cases is the best evidence of what the law is; indeed, it makes law.' . . . I doubt whether today anyone would claim that a uniform practice of the profession makes the law. But in the present context it does have a significant impact. In holding that money paid under a mistake of law is recoverable, an essential factor is that the retention of the money so paid would constitute an unjust enrichment of the payee . . . If, at the date of payment, it was settled law that payment was legally due, I can see nothing unjust in permitting the payee to retain moneys he received at a time when all lawyers skilled in the field would have advised that he was entitled to receive them and the payer was bound to pay them."

182 Cf. Colombi Ciacchi, "Non-legislative harmonisation", n. 53 above.

183 [2002] UKHL 52.

184 Ibid. para. 17.

185 M Kenny and J Devenney, "The fallacy of the common core: polycontextualism in surety protection – a hard case in harmonisation discourse" in M Andenæs and C Andersen (eds), *The Theory and Practice of Harmonisation* (Cheltenham: Edward Elgar 2011).

186 See P O'Callaghan, "Protection from unfair suretyships in Ireland" in Colombi Ciacchi (ed.), *Protection of Non-Professional Sureties*, n. 61 above; J Mee, "Undue influence and bank guarantees" (2002) 37 *Irish Jurist* 292.

187 *OFT* v *Abbey National plc* [2009] UKSC 6, para. 91.

Yet any such challenge would probably, at least by implication, need to challenge the particular cross-subsidy which banks employ in the United Kingdom's largely "free if in credit" banking system; and it is submitted that it would be a brave court that would categorise such an ingrained system as "unfair". The point is, of course, that any EU harmonisation agenda in this area is constrained by, for example, social and cultural norms:

> different benchmarks in member states when reviewing contractual terms . . . Accordingly, traders cannot use a contractual clause which is valid across the EU, but must instead formulate different clauses for each member state. Hence, considerable obstacles to the functioning of the internal market exist. Providers can only perform pre-formulated contracts across borders with considerable transaction costs.[188]

Nor is there any real EU jurisprudence on unfair terms[189] to inform the exercise, meaning that "harmonisation under the Directive is more apparent than real".[190]

Conclusions

As Lord Walker candidly acknowledged,[191] the decision of the Supreme Court in *OFT* v *Abbey National plc* is likely to be a disappointment for many consumers. It is, of course, true that the Supreme Court did not fully shut the door on the possibility of a successful challenge to the relevant charges under the Regulations. Yet – given that any such challenge would probably, at least by implication, need to challenge the particular cross-subsidy which banks employ in the United Kingdom's largely "free if in credit" banking system – the chances of such a challenge being successful seem remote. This underlines the difficulty of harmonising across the EU concepts, such as "unfairness", which are loaded with social, economic and behavioural norms.[192] Indeed, *Abbey* illustrates some of the other challenges for any EU harmonisation agenda. In particular, *Abbey* raises issues in relation to the transposition of EU Directives. As noted above, the Regulations follow very closely the wording of the Directive. Whilst such an approach to the implementation of the Directive might be viewed as ensuring compliance with the obligations of the Directive, it can also, to an extent, be viewed as regrettable for, at least, three inter-related reasons. First, subject to overriding EU obligations, EU Directives may provide a unique opportunity to develop national law in a rational, and non-fragmented, manner so as to meet EU obligations. One may, of course, lament the overlap between the Regulations and the Unfair Contract Terms Act 1977;[193] and one might, encouraged by the views of the Law Commission,[194] wish for one discrete, self-contained piece of legislation on unfair terms. Yet, there is a wider issue here: the method used to transpose the Directive does little to promote coherence in domestic law by, *for example*, seeking to accommodate it within the traditional theoretical dichotomy in English law between procedural and substantive unconscionability[195] (even if that theoretical framework might have needed to be developed for such a purpose).[196]

188 Schulte-Nölke, *EC Consumer Law Compendium*, n. 100 above. Cf. also the German cases on bank charges which are discussed by M Kenny, "Orchesrating sub-prime consumer protection in retail banking: Abbey National in the context of Europeanised private law" (forthcoming 2011) *European Review of Private Law*.

189 See n. 60 above.

190 Report on Directive 93/13/EEC on Unfair Terms in Consumer Contracts (Com (2000) 248 final), p. 30.

191 *OFT* v *Abbey National plc* [2009] UKSC 6, para. 52.

192 See Kenny et al., "Conceptualising unconscionability", n. 175 above.

193 Cf. Peel, *Treitel*, n. 38 above, pp. 290–2.

194 Law Commission, *Unfair Terms in Contracts*, Law Comm. No 292 (2005).

195 See, for example, *Hart* v *O'Connor* [1985] 2 WLR 944, para. 958, per Lord Brightman, discussed in Devenney and Chandler, "Unconscionability", n. 145 above, pp. 544–51.

196 Cf. Teubner, "Legal irritants", n. 57 above.

Secondly, such an approach *may* (as was arguably the case in *Abbey*) encourage, and ostensibly legitimise, a purely literal approach to the interpretation of measures which transpose EU directives. Thirdly, where the relevant directive, as is argued to be the case here, contains essentially an uneasy and somewhat opaque political compromise provision, which needs to be interpreted against complex economic and social considerations, this approach to transposition essentially abdicates the task of cutting through the knot to the judiciary. Yet it is not clear that the judiciary is always best placed to provide the Alexandrian-like solution; and, indeed, to some extent in *Abbey*, Lord Walker[197] and Lady Hale[198] pushed the responsibility back to Parliament.[199]

197 *OFT* v *Abbey National plc* [2009] UKSC 6, para. 50.

198 Ibid. para. 92.

199 See also BBC interview with Vince Cable MP, "Vince Cable: banks continue to rip customers off", available at www.bbc.co.uk/news/uk-10664533. Cf. position of OFT, available at www.oft.gov.uk/OFTwork/markets-work/completed/personal/.

NILQ 62(1): 57–74

We need to talk: "democratic dialogue" and the ongoing saga of prisoner disenfranchisement

C R G MURRAY*

Lecturer, Newcastle Law School

Abstract

Challenges to prisoner disenfranchisement in the United Kingdom have persisted for more than a decade, progressing through the domestic courts to the European Court of Human Rights and back again. The issue has been subject to a prolonged two-stage consultation. And yet, in spite of the decision in Hirst *v* UK (No 2) *that the current disenfranchisement regime breaches the right of prisoners to vote, the governments in office since this decision have to date declined to introduce legislation to rectify the breach. This article considers the response of the United Kingdom's domestic courts to this situation, focusing upon the general unwillingness of the courts to confront the government over the stalled reform process and the implications of this reluctance for the operation of the Human Rights Act 1998 (HRA). The prisoner enfranchisement cases give rise to important questions regarding the domestic courts' discretion to re-interpret provisions so as to bring the law within the margin of appreciation and whether multiple declarations of incompatibility should be issued if the government fails to respond to the first in an appropriate and timely manner.*

Key words: prisoner disenfranchisement; right to vote; Human Rights Act; separation of powers; margin of appreciation

Introduction

The issue of prisoner voting rights in the United Kingdom has remained unresolved since the decision of the European Court of Human Rights' Grand Chamber in *Hirst* v *UK (No 2)*.[1] Hirst, serving a life sentence for voluntary manslaughter, claimed that the denial of his right to vote at the 2001 general election constituted an infringement of Article 3 of Protocol 1 of the European Convention on Human Rights (ECHR), by which states "undertake to hold free elections at reasonable intervals by secret ballot, under conditions which will ensure the free expression of the opinion of the people in the choice of the legislature". Prison officials had refused his requests to be added to the register of electors on the basis of the Representation of the People Act 1983, which provides that "[a] convicted person during the time that he is detained in a penal institution in pursuance of his sentence . . . is legally incapable of voting at any parliamentary or local government election".[2]

* My thanks to Kevin J Brown (Newcastle University), Roger Masterman (Durham University), Richard Mullender (Newcastle University) and Aoife O'Donoghue (Durham University) for their advice and comments upon earlier drafts of this article. Any errors remain my own.

1 *Hirst* v *United Kingdom (No 2)*, App. No 74025/01 (2006) 42 EHRR 41.

2 Representation of the People Act 1983, s. 3(1).

Ultimately, the Grand Chamber recognised that the blanket ban on prisoner voting violated the right to vote, rejecting the United Kingdom government's suggestions that "imprisonment after conviction involves the forfeiture of rights beyond the right to liberty, and especially the assertion that voting is a privilege not a right".[3] The reform process which would ordinarily be set in motion by such a decision was, however, stymied by a pair of prolonged consultations. These consultations did not commence until over a year after the Grand Chamber's decision[4] and only concluded in September 2009.[5] This represents a significantly delayed and, according to Parliament's Joint Committee on Human Rights (JCHR), "disproportionate"[6] consultation process which made a mockery of the Labour government's original assurances that remedial legislation would be introduced by October 2007.[7] By the time of the 2010 general election, prisoners remained unable to vote in the United Kingdom. This situation drew severe criticism from the Council of Europe's Committee of Ministers, which:

> [E]xpressed profound regret that despite the repeated calls of the Committee, the United Kingdom general election was held on 6 May 2010 with the blanket ban on the right of convicted prisoners in custody to vote still in place.[8]

This rebuke calls into question the "hands-off" approach of the United Kingdom's domestic courts during the attenuated reform process. In evaluating their performance, this article firstly considers the Grand Chamber's decision in *Hirst* that the United Kingdom's blanket disenfranchisement of prisoners was incompatible with Article 3 of Protocol 1, which provided the backdrop to the domestic courts' responses to renewed challenges to the voting ban. Faced with the Labour government's protracted consultations, prisoners attempted to enforce the right to vote at various elections held in the United Kingdom (including elections for devolved institutions and European Parliament elections), confronting courts in Scotland,[9] Northern Ireland,[10] and England and Wales[11] with the ongoing breach of their right to vote. The courts' reactions to these challenges showcase the limitations upon the remedial mechanisms established under the HRA where a government prevaricates rather than addressing an identified breach. These cases also betray the discomfort of some judges with suggestions that, under the HRA, the judiciary should engage in a "democratic dialogue" with other branches of government.[12] This discomfort manifested itself in the failure of these courts to address the limitations of the existing reform proposals on prisoner enfranchisement. Therefore, if the HRA is generally regarded as being responsible for a "quantum leap into a new legal culture of fundamental rights and

3 *Hirst (No 2)* (2006) 42 EHRR 41, para. 75.

4 See Consultation Paper CP39/06, *Voting Rights of Convicted Prisoners Detained within the United Kingdom: The UK's response to the Grand Chamber of the European Court of Human Rights judgment in* Hirst *v* United Kingdom, 14 December 2006, available at www.dca.gov.uk/consult/voting-rights/cp2906.pdf.

5 See Consultation Paper CP6/09, *Voting Rights of Convicted Prisoners Detained within the United Kingdom: Second Stage Consultation*, 8 April 2009, www.justice.gov.uk/consultations/docs/prisoner-voting-rights.pdf.

6 JCHR, 16th Report of 2006–07, *Monitoring the Government's Response to Court Judgments Finding Breaches of Human Rights*, 18 June 2007, HL Paper 128/HC 728, para. 78.

7 See *Smith* v *Scott* [2007] CSIH 9, para. 40.

8 CM/Del/Dec (2010) 1086/18, 7 June 2010, 1086th Meeting (DH).

9 *Smith* [2007] CSIH 9 and *Traynor* v *Secretary of State for Scotland* [2007] CSOH 78.

10 *Toner and Walsh* [2007] NIQB 18.

11 *R (Chester)* v *Secretary of State for Justice* [2009] EWHC 2923 (Admin).

12 See R Clayton QC, "Judicial deference and 'democratic dialogue': the legitimacy of judicial intervention under the Human Rights Act 1998" (2004) *PL* 33.

freedoms",[13] the reluctance of the courts to hasten this particular reform process helps us to measure the limits of this jump.

Slow progress towards prisoner enfranchisement in the UK's courts

The now vexed issue of prisoner disenfranchisement was long neglected by the United Kingdom's courts. Reviewing Canadian and American cases on prisoners' rights in the 1970s, Graham Zellick admitted that, by contrast, within the legal systems of the United Kingdom, "the law for most purposes tends to stop at the prison gates, leaving the prisoner to the almost exclusive control of the prison authorities".[14] In the 1960s and 1970s the courts in the United States had to determine a series of cases on voting rights and the rights of offenders. These streams of jurisprudence merged in the case of *Richardson* v *Ramirez*,[15] a challenge to the continuing disenfranchisement of offenders even after their release. Although Justice Rehnquist conceded that "the more enlightened and sensible" approach to criminal justice would be to harness the rehabilitative potential of such an extension of the franchise, the Supreme Court ultimately decided to leave the issue to state legislatures.[16]

Whilst the ongoing disenfranchisement of offenders after their release was therefore pushed up the political agenda in the United States by such legal challenges, the dearth of similar litigation by prisoners in the United Kingdom in this era can largely be explained by the historic restrictions on their access to the courts imposed as part of the ancient concept of "civil death".[17] Even after the restrictions on the ability of prisoners to access the courts were repealed,[18] and despite prisoners gaining access to legal aid to fund litigation,[19] the Prison Rules continued to curtail prisoners' ability to conduct their legal affairs without leave from the Home Secretary.[20] The courts, until the 1980s, remained reluctant to question the operation of the Prison Rules,[21] cowed by Goddard LJ's assertion that:

> It would be fatal to all discipline in prisons if governors and warders had to perform their duty always with the fear of an action before their eyes if they in any way deviated from the rules.[22]

The courts' prolonged failure to address prisoners' rights left the domestic legal systems increasingly out of step with other common law jurisdictions.[23] Pressure on the courts' stance culminated in Lord Wilberforce's acceptance, in *Raymond* v *Honey*,[24] that "a convicted prisoner, in spite of his imprisonment, retains all civil rights which are not taken away

13 W Wade, "Human rights and the judiciary" (1998) *EHRLR* 520, p. 532.

14 G Zellick, "Prisoners' rights in England" (1974) 24 *UTLJ* 331, p. 331.

15 *Richardson* v *Ramirez* (1974) 418 US 24.

16 Ibid. p. 55.

17 For an explanation of the legal disabilities which Parliament imposed on prisoners until the Criminal Justice Act 1948, see the Forfeiture Act 1870, s. 8.

18 See Criminal Justice Act 1948, s. 70(1).

19 See Legal Aid and Advice Act 1949, s. 7. These provisions are now contained within the Access to Justice Act 1999, s.7.

20 Prison Rules 1964 (SI 1964/388), r. 34(8). This situation persisted until a prisoner complained to the European Commission of Human Rights in *Knechtl* v *United Kingdom*, App. No 4115/69, 13 *YB European Convention on Human Rights* 730.

21 See M Loughlin and P Quinn, "Prisons, rules and courts: a study in administrative law" (1993) 56 *MLR* 497, p. 501.

22 *Arbon* v *Anderson* [1943] KB 252, p. 255. In the late 1970s, Shaw LJ echoed this proposition, asserting that under the Prison Act "the courts have no defined place and no direct or immediate function" in *R* v *Board of Visitors of Hull Prison ex parte St Germain* [1979] 1 QB 425, p. 454.

23 See S Easton, "Constructing citizenship: making room for prisoners' rights" (2008) 30 *JSWFL* 127, p. 135.

24 *Raymond* v *Honey* [1983] 1 AC 1.

expressly or by necessary implication".[25] For Susan Easton, this decision and the cases which followed "brought the prisoner and prison life firmly back within the notion of citizenship and into the arena of judicial scrutiny".[26] Nevertheless, governments continued to argue that rights, including the right to conduct privileged legal correspondence[27] and right of access to journalists,[28] became privileges for the duration of an individual's imprisonment. In response, the courts reaffirmed that, whilst imprisonment necessarily entails a curtailment of rights beyond the liberty of the individual, this does not mean that other rights are extinguished.[29]

When, drawing upon this sea change in the judiciary's approach to prisoner rights, three individuals held in English prisons ultimately challenged their disqualification from voting,[30] they argued that the right to vote could not be removed as incidental upon imprisonment. Jack Straw, the then Home Secretary, persisted with the argument that an individual's ability to vote is a function of his or her good citizenship, asserting that detention was but "one element" of punishment and that "[r]emoval from society means removal from the privileges of society, amongst which is the right to vote for one's representative".[31] Given the courts' acceptance since *Raymond* v *Honey* that imprisonment primarily involved a deprivation of liberty, this familiar proposition appeared to place the government on unstable ground. Nonetheless, in *Pearson and Martinez*, the Divisional Court differentiated the right to vote from rights such as access to a journalist, with Kennedy LJ asserting that:

> [There is] no reason why Parliament should not, if so minded, in its dual role as legislator in relation to sentencing and as guardian of its institutions, order that certain consequences shall follow upon conviction or incarceration without transgressing in any way the philosophy expounded in *Raymond* v *Honey*.[32]

The Divisional Court heard *Pearson and Martinez* less than six months after the HRA came into operation. This altered context raised the case's profile. Immediately prior to the hearing, the shadow Home Secretary, Ann Widdecombe, condemned this "darn silly" case as the upshot of permitting individuals to claim that the state had breached the ECHR in the domestic courts.[33] Her comments could also be read as agitating for the rejection of the claim, especially as she skirted the *sub judice* rule by speaking outside Parliament.[34] The Divisional Court's judgment, perhaps with one eye to the controversy surrounding the case, mirrored the US Supreme Court decision in *Richardson* v *Ramirez* that reform was a matter for the legislature.[35] It accepted that:

25 *Raymond* v *Honey* [1983] 1 AC 1, p. 10. In making this assertion Lord Wilberforce acknowledged the influence of the Canadian Supreme Court's decision in *Solosky* v *The Queen* (1979) 105 DLR (3d) 745.

26 Easton, "Constructing Citizenship", n. 23 above, p. 136.

27 *R* v *Secretary of State for the Home Department ex parte Leech (No 2)* [1994] QB 198.

28 *R* v *Secretary of State for the Home Department ex parte Simms* [2000] 2 AC 115.

29 See *ex parte Leech* [1994] QB 198, p. 209, per Steyn LJ. See also *ex parte Simms* [2000] 2 AC 115, p. 143, per Lord Hobhouse.

30 *R (Pearson and Martinez)* v *Secretary of State for the Home Department* [2001] EWHC 239 (Admin).

31 Ibid. para. 9.

32 Ibid. para. 18.

33 BBC News, 21 March 2001, http://news.bbc.co.uk/1/hi/uk/1233084.stm.

34 It should be noted that, under the *sub judice* rule in force at the time, a civil case could be discussed by MPs in Parliament where it related to issues of national importance. See Select Committee on Procedure, *Matters Sub Judice*, 4th Report of Session 1971–72, 27 June 1972, HC 298.

35 *Richardson* (1974) 418 US 24.

> [T]here is a broad spectrum of approaches among democratic societies [to prisoner disenfranchisement], and the United Kingdom falls into the middle of the spectrum . . . its position in the spectrum is plainly a matter for Parliament not for the courts.[36]

Hirst v *UK*: the breach of the right to vote

The third claimant in *Pearson and Martinez*, John Hirst, remained unconvinced that his desire to vote was "darn silly". He successfully pursued his case before the European Court of Human Rights.[37] The Labour government responded to this setback by requesting that the case be referred to the Grand Chamber,[38] arguing once more that the punishment of being banned from voting had the effect of "enhancing civic responsibility and respect for the rule of law by depriving those who had breached the basic rules of society of the right to have a say in the way such rules were made for the duration of their sentence".[39] Hirst countered that the blanket ban was disproportionate, being "unrelated to the nature or seriousness of the offence and varied in its effects on prisoners depending on whether their imprisonment coincided with an election"[40] and that, contrary to the government's assertions, it "took away civic responsibility and eroded respect for the rule of law, serving to alienate prisoners further from society".[41]

Accepting that automatic disenfranchisement infringed the right to vote, the Grand Chamber evidently considered that the Labour government's stance was predicated on a concern for the general public's perceived antipathy towards prisoner enfranchisement:

> There is . . . no question that a prisoner forfeits his Convention rights merely because of his status as a person detained following conviction. Nor is there any place under the Convention system, where tolerance and broadmindedness are the acknowledged hallmarks of democratic society, for automatic disenfranchisement based purely on what might offend public opinion.[42]

As Steve Foster acknowledges, however, the Grand Chamber did not reject all voting restrictions on prisoners. Rather, "it accepted, at least in principle, that the right to vote may be lost as part of an individual sentence".[43] The Grand Chamber refused to criticise the United Kingdom's argument that removal of the vote from individuals in response to their criminal conviction had the potential to encourage good citizenship. It asserted that there was "no reason in the circumstances of this application to exclude these aims as untenable or *per se* incompatible with the right [to vote]".[44] Instead, the majority emphasised that it was the blanket nature of the disenfranchisement imposed by the Representation of the People Act 1983 which breached Article 3 of Protocol 1, on the basis that "the principle of proportionality requires a discernible and sufficient link between the sanction and the conduct and circumstances of the individual concerned".[45] These statements indicate the interplay between proportionality and margin of appreciation in *Hirst*. Building upon the wide margin of appreciation in this field, the Grand Chamber's approach recognised the

36 *Pearson and Martinez* [2001] EWHC Admin, para. 41.
37 *Hirst* v *United Kingdom (No 2)* (2004) 38 EHRR 40.
38 *Hirst (No 2)* (2006) 42 EHRR 41.
39 Ibid. para. 50.
40 Ibid. para. 45.
41 Ibid. para. 44.
42 Ibid. para. 70.
43 S Foster, "Reluctantly restoring rights: responding to the prisoner's right to vote" (2009) 9 *HRLR* 489, p. 498.
44 *Hirst (No 2)* (2006) 42 EHRR 41, para. 75.
45 Ibid. para. 71.

disproportionate nature of a blanket ban but allowed states to deny the vote to particular prisoners on the basis of the seriousness or nature of their offences.

In the aftermath of *Hirst*, the extent of the leeway provided to states in the Grand Chamber's decision was not widely recognised. The decision was portrayed by opponents of the ban on prisoners voting as a fatal blow to the United Kingdom's disenfranchisement regime. This was certainly how *Hirst* was received in Ireland, where legislation was quickly enacted to end the disenfranchisement of prisoners.[46] Piloting the Electoral (Amendment) Bill 2006 through the Dáil, Minister for the Environment, Heritage and Local Government, Dick Roche, could not resist highlighting the delays in the United Kingdom's legislative response to *Hirst*: "It is consulting; we are acting."[47] As recently as October 2008, Parliament's JCHR claimed that *Hirst* provided "clear guidance that individuals' fundamental human rights, including the right to vote, are not contingent on their continuing to be 'good citizens'".[48]

This interpretation did not reflect the nuance inherent in the *Hirst* decision. By March 2010, however, the Labour government had persuaded the JCHR to recognise that "the Grand Chamber left a broad discretion to the United Kingdom to determine how to remove the blanket ban".[49] The significance of this volte-face should not be underestimated; *Hirst* was not the final word on prisoner voting rights. In the context of the United States, Andrew Shapiro has described the Supreme Court's refusal to rule that felon disenfranchisement was unconstitutional in *Richardson* v *Ramirez*[50] as condemning opponents of criminal disenfranchisement to a long-running battle through the state legislatures, achieving only "piecemeal, incremental reform".[51] Similarly, the flexibility of the Grand Chamber's decision in *Hirst* returned the focus of the debate over enfranchisement to the national institutions of states party to the ECHR.

After *Hirst*: consultations and renewed legal challenges

Using the full extent of the room for manoeuvre provided by the *Hirst* decision, the Labour government's response to the issue of prisoner enfranchisement was conditioned by the political imperative not to be seen as "soft on crime".[52] Under the premierships of both Tony Blair and Gordon Brown, the Department for Constitutional Affairs and its successor, the Ministry of Justice issued consultation documents on reform (in December 2006 and April 2009) which suggested that the government favoured, at most, a very limited

46 For a summary of the debates leading to the enactment of the Electoral (Amendment) Act 2006 in Ireland, see C Behan and I O'Donnell, "Prisoners, politics and the polls" (2008) 48 *Brit J Criminol* 319, pp. 326–8.

47 D Roche, TD, Dáil Éireann Debates, vol. 624, col. 1988, 5 October 2006.

48 JCHR, 31st Report of 2007–08, *Monitoring the Government's Response to Human Rights Judgments: Annual Report 2008*, 7 October 2008, HL Paper 173/HC 1078, para. 58.

49 JCHR, 15th Report of Session 2009–10, *Enhancing Parliament's Role in Relation to Human Rights Judgments*, 9 March 2010, HL Paper 85/HC 455. para. 107.

50 *Richardson* (1974) 418 US 24.

51 A Shapiro, "Challenging criminal disenfranchisement under the Voting Rights Act: a new strategy" (1993) 103 *Yale LJ* 537, p. 564. Despite these reforms, other writers have tracked rapid increases in the number of individuals disenfranchised under these provisions in the United States from the 1970s onwards. See C Uggen and J Manza, "Democratic contraction? The political consequences of felon disenfranchisement in the United States" (2002) 67 *American Sociological Review* 777.

52 During the 2010 general election, many Labour candidates, particularly those facing strong Liberal Democrat challenges, campaigned on the government's refusal to extend the vote to prisoners. The national Labour Party even had to order one of its candidates to withdraw a leaflet attacking the Liberal Democrat policy of legislating to enfranchise some prisoners on the basis of *Hirst*. The leaflet asserted: "Do you want convicted murderers, rapists and paedophiles to be given the vote? The Lib Dems do": BBC News, 19 April 2010, available at http://news.bbc.co.uk/1/hi/uk_politics/election_2010/8630001.stm.

enfranchisement of prisoners. These consultations proceeded on the basis that "the more serious the offence that has been committed, the less right an individual should have to retain the right to vote when sentenced to imprisonment".[53] The documents proposed that franchise restrictions would apply after a low threshold sentence and established that the government would not countenance granting the vote to prisoners serving sentences of more than four years' imprisonment.[54] In Steve Foster's opinion, these proposals were "aimed at allowing the government to encroach upon the prisoner's right to vote to the maximum extent that might be allowed by the Convention and the European Court".[55]

Progress towards enacting even these limited reforms was slow. After Strasbourg's first decision in *Hirst*,[56] Juliet Lyon, Director of the Prison Reform Trust, expressed her concern that the Labour government would seek to delay the enfranchisement of prisoners until after the 2005 general election.[57] Five years later an exasperated JCHR asserted that the obstacle blocking the reform process was simply that the issue of prisoner enfranchisement was "legally straightforward, but politically difficult".[58] These delays returned the incompatibility of the disenfranchisement of prisoners under the ECHR to the arena of the domestic courts. Nonetheless, following *Hirst*, prisoners might have hoped that domestic judges would finally begin the task of "clearing the channels" of obstacles impeding the operation of the democratic process,[59] rather than perpetuating the deference historically shown to Parliament regarding voting rights.

The first case to arise in the wake of *Hirst* occurred as a result of the refusal of an electoral registration officer to include William Smith, a prisoner held in HMP Glenochil, on the electoral register for the Scottish Parliament elections held in May 2003. Smith's appeal against this decision was not heard until late 2006. Three Court of Session judges refused to "read down" s. 3 of the Representation of the People Act 1983 to conform with the right to vote,[60] but did issue a declaration of incompatibility recognising that this provision breached the ECHR.[61] Soon after *Smith* came *Toner and Walsh*,[62] an action bought by two prisoners held in Northern Ireland prior to the Northern Ireland Assembly elections held in March 2007. Given the Grand Chamber's recognition that the blanket disenfranchisement of prisoners breached Article 3 of Protocol 1 and the recent declaration of incompatibility issued regarding the Representation of the People Act in *Smith*, the claimants' challenge focused instead on invalidating Article 4 of the Northern Ireland Assembly (Elections) Order 2001. This secondary legislation set out that the franchise for Northern Ireland Assembly elections mirrored that for local elections, thereby incorporating the ban on prisoner voting. However, Gillen J recognised that the root of the problem remained the Representation of the People Act and that it would be "invidious to

53 Consultation Paper CP6/09, *Second Stage Consultation*, n. 5 above, p. 24.

54 Ibid.

55 S Foster, "Prisoners' rights under threat: do prisoners retain their democratic rights?" (2007) 12 *Cov LJ* 42, p. 49.

56 *Hirst (No 2)* (2004) 38 EHRR 40.

57 See M Dhami, "Prisoner disenfranchisement policy: a threat to democracy?" (2005) 5 *Analyses of Social Issues and Public Policy* 235, p. 238.

58 JCHR, 31st Report, n. 48 above, para. 62.

59 J H Ely, *Democracy and Distrust* (Cambridge, MA: Harvard University Press 1980), p. 105.

60 See *Smith* [2007] CSIH 9, para. 27.

61 Ibid. para. 56.

62 *Toner and Walsh* [2007] NIQB 18.

invalidate an Order that would serve to disenfranchise the voting population of Northern Ireland in circumstances where the genesis of the problem manifestly lies elsewhere".[63]

In *Toner and Walsh*, the delicate position of the peace process in Northern Ireland in the wake of the St Andrews' Agreement,[64] with the pending election potentially "an integral part of that progress",[65] hung over the challenge to prisoner disenfranchisement. In this context, Gillen J unsurprisingly refused to "take a risk laden step which would imperil or prejudice the forthcoming election at such short notice [and] would visit hardship and detriment on the concept of good administration in Northern Ireland".[66] No such impediments weighed on the mind of Burton J in *Chester*.[67] This case involved a claimant who, like John Hirst, remained imprisoned in spite of completing the minimum-term element of his life sentence.[68] Peter Chester sought to register for the 2009 European Parliament elections and claimed that his right to vote had been infringed after permission to do so was denied. Burton J, giving the strong impression that he regarded this claim as a dubious use of legal aid funding,[69] refused to employ either s. 3 or s. 4 of the HRA in response to the breach of the right to vote.[70] In all of these cases, the government accepted that the ban on voting, imposed on all convicted criminals serving custodial sentences, breached the right to vote. The reluctance of the courts to tackle this breach, which in *Chester* manifested itself fully four years after the Grand Chamber's ruling in *Hirst*, leaves an obvious question mark hanging over the effectiveness of the HRA's remedial regime.

Judicial aversion towards the HRA's re-interpretive power

The reluctance of the courts hearing prisoner enfranchisement to use s. 3 of the HRA to create exceptions to the ban on prisoner voting under the Representation of the People Act 1983 arguably indicates a continued reticence amongst judges towards the rights of prisoners. This HRA provision requires that, "[s]o far as it is possible to do so, primary legislation and subordinate legislation must be read and given effect in a way which is compatible with the Convention rights".[71] In *Smith*, the Registration Appeal Court of Scotland asserted that the blanket nature of the ban precluded the use of the re-interpretation provision as there was "no 'grain of the legislation' which could properly serve as a starting point for any interpretation designed to clothe some or all of such prisoners with voting rights".[72] Following the reasoning of Lord Nicholls in *Re S (Minors)*,[73] the court considered that re-interpreting s. 3 of the Representation of the People Act so as to allow any prisoners the vote would "depart substantially from a fundamental feature of the legislation" and would amount to the court "legislating on its own account".[74] Such refusals to caveat or read down these provisions, on the basis that

63 *Toner and Walsh* [2007] NIQB 18, para. 9(viii).

64 This agreement, published on 13 October 2006, was the culmination of negotiations to restore the devolved institutions within Northern Ireland, which had been suspended since 14 October 2002. The full text of the Agreement is available at http://web.archive.org/web/20080610001151/www.standrewsagreement.org/agreement.htm.

65 *Toner and Walsh* [2007] NIQB 18, para. 9(x).

66 Ibid.

67 *Chester* [2009] EWHC 2923 (Admin).

68 Criminal Justice Act 2003, Sch. 22, para. 3.

69 See *Chester* [2009] EWHC 2923 (Admin), para. 19.

70 Ibid. paras 31 and 44.

71 HRA, s. 3(1).

72 *Smith* [2007] CSIH 9, para. 26. See also *Ghaidan* v *Godin-Mendoza* [2004] UKHL 30, para. 33, per Lord Nicholls.

73 *Re S (Minors) (Care Order: Implementation of Care Plan)* [2002] UKHL 10, para. 40.

74 *Smith* [2007] CSIH 9, para. 27.

doing so would amount to a fundamental alteration of the statute, are common across the prisoner enfranchisement cases.[75]

This reliance on the decision in *Re S (Minors)* was, however, misplaced. It neglected the emphasis which the Grand Chamber in *Hirst* placed on the lack of parliamentary debate concerning the purpose of the voting ban (either when the modern formulation of the ban was adopted in the Representation of the People Act 1969, or since).[76] This lack of reasoning behind the imposition of the ban undermines the contention that stipulating specific exceptional categories of prisoners entitled to vote conflicts with any fundamental statutory purpose. As David Feldman noted before the HRA even entered force, the courts' new interpretative power could be usefully employed to place the onus on Parliament to express its legislative intention clearly.[77] Put plainly, if Parliament fails to articulate the reasoning behind a measure (whether it is enacted before or after the HRA), the courts should be able to assess its core purpose and which features of the measure are fundamental to that purpose.

Furthermore, particular groups of prisoners can be separated from the prison population as a whole. Peter Chester had served the minimum-term element of his sentence and was therefore being held not as a punishment, but on grounds of the risk which he posed to the public.[78] Burton J rejected efforts to persuade him to address the particular issue of this category of prisoners. Succumbing to what Joseph Raz identified as the traditional judicial animus towards "partial reform",[79] Burton J considered that a re-interpretation which sought to exclude this group of prisoners from the voting ban, on the basis that removal of the vote was a punishment, would encourage further claims, leading "to piecemeal and possibly continuous amendments, without consideration by Parliament, of legislation dealing with matters of important social policy, all depending upon which claimant happened to be before the Court at any one time".[80] Locating this reasoning within the approach mapped out by Lord Nicholls in *Re S (Minors)*, Burton J concluded that the potential for future litigation amounted to an "important practical repercussion" militating against re-interpretation.[81] Similarly, in *Smith*, the Registration Appeal Court accepted that:

> [T]here were many possible levels at which the line might be drawn for the enfranchisement or disenfranchisement of convicted prisoners in different categories, and it could be no part of this court's function to make an uninformed choice among such alternatives.[82]

This reasoning also consciously echoes Lord Hobhouse's speech in *Bellinger* v *Bellinger*,[83] concerning whether the courts should employ s. 3 of the HRA to re-interpret "female", under s. 11(c) of the Matrimonial Clauses Act 1973, to include a transsexual woman. Rejecting this course of action, Lord Hobhouse declared that it would involve "making a legislative choice as to what precise amendment was appropriate".[84] The context of this

75 See *Chester* [2009] EWHC 2923 (Admin), para. 26.

76 See *Hirst (No 2)* (2006) 42 EHRR 41, para. 79.

77 See D Feldman, "The Human Rights Act 1998 and constitutional principles" (1999) 19 *LS* 165, p. 186.

78 Law Commission, *Murder, Manslaughter and Infanticide*, Law Comm No. 304 (2006), para. 1.58.

79 J Raz, *The Authority of Law* (Oxford: OUP 1979), p. 201.

80 *Chester* [2009] EWHC 2923 (Admin), para. 30.

81 *Re S* [2002] UKHL 10, para. 40. See also *Re P (Adoption: Unmarried Couple)* [2008] UKHL 38, para. 55, per Lord Hope.

82 *Smith* [2007] CSIH 9, para. 26.

83 *Bellinger* v *Bellinger* [2003] UKHL 21.

84 Ibid. para. 78.

decision is, however, significant. The case of *Goodwin* v *UK*,[85] in which Strasbourg concluded that the Matrimonial Clauses Act 1973 breached Articles 8 and 12 ECHR, was decided only six months prior to the House of Lords' hearing in *Bellinger*. The government had accepted the decision in *Goodwin* and announced that it was taking steps to rectify the breach through legislation even before the House of Lords heard *Bellinger*. Even in *Smith*, decided just over a year after the Grand Chamber's decision in *Hirst*, the Registration Appeal Court drew attention to how quickly Parliament had reacted to *Goodwin* by comparison to *Hirst*.[86]

The backdrops to *Bellinger* and *Chester* are therefore markedly different. *Chester* was decided fully four years after *Hirst*. This delay may, of itself, have warranted a more purposive judicial approach to the Representation of the People Act 1983, despite the practical ramifications alluded to by Burton J. As Mary Arden has pointed out in her extra-judicial writing, even after *Re S* and *Bellinger* the courts are still "feeling their way towards a set of rules and canons of construction that will apply where section 3(1) is in point".[87] *Chester* represents a missed opportunity to refine the approach of the courts to the re-interpretive power. The ability of courts to re-interpret a provision "so far as it is possible to do so" is arguably context specific. The courts should therefore take account of factors such as the duration of the identified breach in deciding whether to re-interpret a provision. At the very least, the four years of prevarication over reform that preceded *Chester* should have alerted Burton J to the danger that his confident assertion that the provisions imposing the voting ban "are imminently to be considered by the Legislature"[88] would prove a hostage to fortune.

The "mirror" principle and the HRA's re-interpretive power

Burton J went so far as to assert in *Chester* that, even if he did consider s. 3 of the Representation of the People Act 1983 to be capable of re-interpretation on the basis of *Hirst*, he would nonetheless refuse to do so.[89] Nor is he alone in this assertion, for the domestic courts assessing *Hirst* have uniformly rejected arguments that the Grand Chamber's decision mandated particular reforms. This unwillingness to countenance a judicial solution to the issue of prisoner disenfranchisement is therefore an unexpected side effect of Strasbourg's readiness in *Hirst* to grant leeway to the United Kingdom as to how to reform the voting ban. As Gillen J asserted in *Toner and Walsh*:

> I am not persuaded . . . that the United Kingdom Government could and will only find the approval of the European Court in circumstances where disenfranchisement was imposed by a sentencing judge . . . [I]t is clear to me that insofar as the court ventured onto this plane, their comments were non-prescriptive and purely advisory in character.[90]

Such advisory comments are not, of themselves, unusual. Strasbourg rarely mandates specific reforms on the finding of a breach of a particular human right,[91] particularly in

85 *Goodwin* v *UK* (2002) 35 EHRR 447.

86 *Smith* [2007] CSIH 9, para. 54.

87 M Arden, "The interpretation of UK domestic legislation in the light of European Convention on Human Rights jurisprudence" (2004) 25 *Stat LR* 165, p. 179.

88 *Chester* [2009] EWHC 2923 (Admin), para. 29.

89 Ibid. para. 30.

90 *Toner and Walsh* [2007] NIQB 18, para. 9(iii).

91 See, for example, *Dickson* v *United Kingdom* (2008) 46 EHRR 41, para. 89, where Strasbourg's Grand Chamber rejected an applicant prisoner's request to direct the United Kingdom to urgently reconsider a decision not to grant him access to artificial insemination facilities, on the basis that "[t]he Court's function is, in principle, to rule on the compatibility with the Convention of the existing measures and it does not consider it appropriate in the present case to issue the requested direction".

areas where the states enjoy a wide margin of appreciation.[92] But mere "advice" from Strasbourg does create a problem for the operation of the HRA, and particularly the requirement under s. 2(1) that the courts must "take into account" judgments of the European Court of Human Rights in reaching their decisions. In the prisoner enfranchisement cases, this requirement was read in light of Lord Bingham's exhortation in *Ullah*[93] that the courts should do no more than "keep pace with the Strasbourg jurisprudence as it evolves".[94] This approach has become known as the "mirror" principle, as it requires that the content of the rights incorporated under the HRA reflects the content of the ECHR rights in Strasbourg jurisprudence.[95] Constrained by this principle, the domestic courts hearing the prisoner enfranchisement cases considered themselves unable to engage with the recommendations contained in *Hirst*, on the basis that acting upon advice, as opposed to "clear and constant" Strasbourg jurisprudence,[96] risked overrunning their mandate.[97]

Washing his hands of the issue of prisoner disenfranchisement, Burton J quoted directly from the decision in *Hirst*. Strasbourg, he said, had sought to "leav[e] it to the legislature to decide on the choice of means for securing the rights guaranteed by [the ECHR]".[98] The reference to the legislature in this passage should, however, be read as an unfortunate slip by the majority in *Hirst*. Strasbourg's jurisprudence has never recognised that the process of bringing a law within the margin of appreciation is solely a matter for a national legislature. Rather, in cases dating back to *Handyside* v *United Kingdom*,[99] Strasbourg has asserted that the margin of appreciation "is given both to the domestic legislator . . . and to the bodies, judicial amongst others, that are called upon to interpret and apply the laws in force".[100]

Strasbourg's invocation of the margin of appreciation does not, therefore, simply provide domestic legislatures with scope for action, but also serves as an invitation to the domestic courts to consider particular measures in light of the purposes of the ECHR. Whilst the courts can decline such invitations, perhaps on the basis of deference to elected decision makers,[101] doing so does not call into question their ability to act. Burton J's eagerness to leave this matter to the sole province of legislature, in spite of the delays in enacting reforms following *Hirs*t, may nonetheless appear to be in line with the historic deference of the domestic courts regarding voting rights. Indeed, this may have been an understandable approach, had counsel not directed him to Lord Hoffmann's speech in the adoption case of *Re P*.[102] In this case the House of Lords sought to refine the obligation to remain in line with Strasbourg jurisprudence, with Lord Hoffmann emphasising that:

> When [the European Court of Human Rights] says that a question is within the margin of appreciation of a Member State, it is not saying that the decision must

92 See *Hirst (No 2)* (2006) 42 EHRR 41, paras 61 and 84.

93 *R (Ullah)* v *Special Adjudicator* [2004] UKHL 26.

94 Ibid. para. 20.

95 J Lewis, "The European ceiling on human rights" [2007] *PL* 720, p. 720.

96 To use the expression adopted by Lord Slynn in *R (Alconbury Developments Ltd)* v *Secretary of State for the Environment, Transport and the Regions* [2001] UKHL 23, para. 26.

97 It should be noted that addressing Strasbourg jurisprudence in these cases would not have brought these courts into conflict with the principles of *stare decisis*. See *Kay and others* v *London Borough of Lambeth and others* [2006] UKHL 10, para. 43, per Lord Bingham.

98 *Chester* [2009] EWHC 2923 (Admin), para. 50. See also *Hirst (No 2)* (2006) 42 EHRR 41, para. 84.

99 *Handyside* v *United Kingdom* (1976) 1 EHRR 737.

100 Ibid. para. 48.

101 See *R v DPP ex parte Kebilene* [2000] 2 AC 326, per Lord Hope, p. 381.

102 *In re P (and others)* [2008] UKHL 38.

> be made by the legislature, the executive or the judiciary. That is a matter for the Member State.[103]

Lord Hoffmann identified the task of developing the law within the margin of appreciation as one basis upon which the courts could depart from the "mirror" principle in appropriate cases.[104] The courts hearing the prisoner enfranchisement cases, as much as the legislature, therefore had a role to play in bringing domestic law within the margin of appreciation. Instead, the pervasive influence of this principle persuaded these courts that it would not be possible to employ the HRA's re-interpretive power. These cases indicate that strict adherence to the rule in *Ullah* remains the safe option for High Court judges seeking to avoid accusations that they are imposing novel burdens on public authorities, even as the appellate courts move away from the rigidity of the "mirror" principle.[105]

The prisoner enfranchisement cases therefore highlight the tension between the perceived functions of Strasbourg and the domestic courts. Strasbourg conceives its role as identifying breaches of the ECHR rights and that, thereafter, the onus passes to any relevant institutions of the state in breach to bring its laws into conformity with the ECHR.[106] The prisoner enfranchisement cases, by contrast, indicate that many domestic judges remain reluctant to employ their powers under s. 3 of the HRA unless Strasbourg has clearly established the minimum contents of a right. These cases indicate, that even in instances, such as *Hirst*, where Strasbourg has provided advice as to how to bring domestic law in line with ECHR requirements, domestic judges will often seek refuge in a strict conception of the "mirror" principle to avoid having to re-interpret statutory provisions in line with such advice.

Declarations of incompatibility and the separation of powers

Even in light of the above analysis, re-interpreting the Representation of the People Act 1983 to extend the vote to exceptional categories of prisoner would have been a radical step,[107] although one arguably commensurate with the importance of the right to vote. The question of whether to issue a declaration of incompatibility under s. 4 of the HRA, however, should have been much less controversial. In *Bellinger*, despite the government's commitment to rectify the law in light of *Goodwin*, Lord Nicholls concluded that the House of Lords should use its discretion to issue a declaration and "formally record that the present state of statute law is incompatible with the Convention".[108] The court in *Smith* followed this lead, asserting that Lord Nicholls' reasoning "applies equally, or with greater force, to this case", and declared s. 3 of the Representation of the People Act 1983 incompatible with the ECHR.[109] In spite of the government's assurances that it accepted and was taking steps to rectify the law in light of *Hirst*,[110] the court considered it necessary

103 *In re P (and others)* [2008] UKHL 38, para. 32.

104 Ibid, para. 31.

105 In addition to cases such as *In re P and others* [2008] UKHL 38 and *R* v *Horncastle and others* [2009] UKSC 14, in which the House of Lords and Supreme Court qualified the "mirror" principle, Baroness Hale adopted a flexible conception of the reasoning underpinning the principle in Gentle: "[The courts] are not free to foist upon Parliament or upon public authorities an interpretation of a Convention right *which goes way beyond anything which we can reasonably foresee that Strasbourg might do*": *R (Gentle and another)* v *Prime Minister and others* [2008] UKHL 20, para. 56 (emphasis added).

106 See N Krisch, "The open architecture of European Human Rights law" (2008) 71 *MLR* 183, pp. 213–15.

107 See A Kavanagh, *Constitutional Review under the UK Human Rights Act* (Cambridge: CUP 2009), pp. 137–42.

108 *Bellinger* [2003] UKHL 21, para. 55.

109 *Smith* [2007] CSIH 9, paras 54–6.

110 Ibid. para. 51.

to issue the declaration given that "the timetable [for reform] has indeed slipped, and slipped badly".[111]

Still the Labour government did not act. By the time of the hearing in *Chester* the timetable for reform had slipped by a further two years. Burton J, confronted with what he regarded as the exhaustion of the system of remedies established in the HRA, considered that there was nothing more that the courts could do in these circumstances. He refused to use his discretion to issue a further declaration of incompatibility, on the basis that "there is no need for any declaration to be made by yet another court, as one has already been made which is binding on the UK Government".[112] His characterisation of the declaration issued in *Smith* as "binding" upon the government is, however, open to question. As Gavin Phillipson asserts, a government's use of its power under the HRA to amend legislation[113] to rectify a declared incompatibility "is not and cannot be guaranteed".[114] Indeed, the Labour government had already expressed its intention not to fast-track reforms to the voting ban under s. 10(2) of the HRA in the wake of *Smith*.[115] Burton J's description of the existing declaration as binding was nonetheless a useful fiction. He asserted that, because a "binding" first declaration had been made, exercising his discretion to issue a second declaration would not simply be redundant but contrary to the separation of powers:

> Insofar as some pressure might have been intended to be brought to bear on the First Defendant . . . to act speedily in bringing forward legislation to Parliament, that would be . . . offensive to constitutional principles.[116]

Burton J's refusal to issue a declaration on the basis that it would jeopardise the separation of powers, while neglecting to criticise the government's failure to rectify such a long-standing breach of human rights, reveals his limited conception of the courts' role since the enactment of the HRA. Academics have frequently mapped out an extensive role for declarations of incompatibility.[117] Aileen Kavanagh, for example, trumpets the ability of declarations to "grab the headlines"[118] and thereby draw public attention to abuses of human rights, whilst Richard Clayton identifies them as facilitating "democratic dialogue" between the branches of government regarding human rights.[119] In the opposite camp, however, the former Court of Appeal judge Richard Buxton adopts a more mechanistic view of declarations as "principally a means of securing the reconciliation of English statute law with the requirements of the ECHR, rather than as a forensic instrument for use in a particular case".[120] Buxton's view won through in *Chester*, underscoring the opposition

111 *Smith* [2007] CSIH 9, para. 43.

112 *Chester* [2009] EWHC 2923 (Admin), para. 34.

113 HRA, s. 10(2).

114 G Phillipson, "Deference, discretion, and democracy in the Human Rights Act era" (2007) 60 *CLP* 40, p. 67. Richard Buxton similarly concludes that s. 4 "envisages, though it does not mandate, remedial action", R Buxton, "The future of declarations of incompatibility" (2010) *PL* 213, p. 221. See also William Wade's criticism that the "executive discretion" as to how to respond to a declaration of incompatibility is "contrary to the rule of law in its most basic sense". Wade, "Human rights", n. 13 above, p. 531.

115 *Chester* [2009] EWHC 2923 (Admin), para. 13.

116 Ibid. para. 34.

117 For Fiona de Londras and Fergal Davis, "[t]he declaration of incompatibility should be seen as a 'moment' capable of galvanizing parliamentary and popular dissent". F de Londras and F Davis, "Controlling the executive in times of terrorism: competing perspectives on effective oversight mechanisms" (2010) 30 *OJLS* 19, p. 46.

118 Kavanagh, *Constitutional Review*, n. 107 above, p. 230.

119 Clayton, "Judicial deference", n. 12 above, pp. 45–6. See also F Klug, "The Human Rights Act – a 'third way' or a 'third wave' Bill of Rights" (2001) *EHRLR* 361, p. 370.

120 Buxton, "The future of declarations of incompatibility", n. 114 above, p. 214.

of some judges to attempts by academic lawyers to conceptualise their role under the HRA as "a form of privileged pressure group whose function it is to raise good reasons why a litigant's interests should be respected".[121] Given the institutional inertia inherent in this mind-set, it is unsurprising that some judges have sought to generate constitutional difficulties with the use of their powers under the HRA.[122]

The *Chester* decision indicates both the discretionary nature of declarations of incompatibility and the fact that declarations, where issued, are not enforceable. These factors will continue to prevent the European Court of Human Rights from accepting that declarations constitute an effective remedy.[123] In *Burden* v *United Kingdom*, the Grand Chamber acknowledged that:

> [I]t cannot be excluded that at some time in the future the practice of giving effect to the national courts' declarations of incompatibility by amendment of the legislation is so certain as to indicate that s. 4 of the Human Rights Act is to be interpreted as imposing a binding obligation.[124]

However, at that time, due to the "relatively small number of such declarations that have become final",[125] Strasbourg declined to recognise s. 4 as providing an effective remedy. Such reticence would appear to be vindicated by the saga of the prisoner enfranchisement cases. These cases affirm Conor Gearty's view that s. 4 "amounts to little more than a cry for action".[126] Worse, if Burton J's refusal to use his discretion to issue a second declaration is followed by other courts, the judiciary would be limited to only ever crying out once in response to a breach of human rights, no matter how long ministers procrastinate over reform of the law.

"Parliament is on the job": the courts and reform proposals

The previous section established how a mechanistic conception of s. 4 of the HRA, as no more than a tool for identifying infringements of rights, can transform requests for multiple declarations of incompatibility into a threat to the separation of powers. This, however, is only one example of the reluctance of courts to be drawn into "democratic dialogue" in the realm of prisoner disenfranchisement. The courts hearing these cases also declined to assess the compatibility of the Labour government's reform proposals with Article 3 of Protocol 1. Beyond general criticism of the pace of reform, such as Gillen J finding "little evidence of a determination to prioritise appropriately the task that was defined by the *Hirst* decision",[127] the domestic courts hearing prisoner disenfranchisement since *Hirst* have largely relied on these proposals to counter the claims of individual prisoners. In spite of the second consultation document's recognition that "ultimately Parliament must debate and decide on the extent of the franchise",[128] the government's representatives in *Chester*[129] (and, prior to

121 T. Hickman, "Constitutional dialogue, constitutional theories and the Human Rights Act 1998" (2005) *PL* 306, p. 309.

122 On the broader issue of the readiness of some courts, practitioners and commentators to identify constitutional difficulties with the role of the courts under the HRA, see Baroness Hale, speech to the Salford Human Rights Conference 2010, 4 June 2010, p. 13, available at www.supremecourt.gov.uk/docs/speech_100604.pdf.

123 Article 35(1) ECHR.

124 *Burden* v *United Kingdom* (2008) 47 EHRR 38, para. 43.

125 Ibid. para. 41.

126 C Gearty, "Revisiting section 3(1) of the Human Rights Act" (2003) 119 *LQR* 551, p. 552.

127 *Toner and Walsh* [2007] NIQB 18, para. 9(ii).

128 Consultation Paper CP6/09, *Second Stage Consultation,* n. 5 above, p. 21.

129 See *Chester* [2009] EWHC 2923 (Admin), para. 14.

the second consultation, in *Toner and Walsh*)[130] emphasised that the claimant prisoners would not be enfranchised under proposals which would only extend the vote to individuals imprisoned for short periods.[131] In *Toner and Walsh*, where the claimants were convicted of armed robbery and burglary respectively, Gillen J asserted that he was "singularly unconvinced that the applicants . . . will ever be able to lay claim to a right to vote".[132]

In contrast to this willingness to consider the proposals where they supported the government's case, in *Chester* the High Court refused to address the inadequacies of the reform proposals, on the basis of the constitutional convention that the courts would not ordinarily comment on proposals before Parliament.[133] This convention operates alongside Article 9 of the Bill of Rights 1689, which provides that "the freedom of speech and debates or proceedings in Parlyament ought not to be impeached or questioned in any court or place out of Parlyament". Several judges have justified the convention as upholding good relations between the courts and Parliament.[134] In *ex parte Smedley*,[135] Sir John Donaldson MR followed his famous statement that "it is a constitutional convention of the highest importance that the legislature and the judicature are separate and independent of one another"[136] by declaring that one facet of convention prevented the courts from discussing proposals which were before Parliament:

> [I]t would clearly be a breach of the constitutional conventions for this court, or any court, to express a view, let alone take any action, concerning the decision to lay this draft Order in Council before Parliament or concerning the wisdom or otherwise of Parliament approving that draft.[137]

Emphasising the importance of this convention in maintaining the separation of powers,[138] Burton J concluded that the proposed reform "plainly involves a balancing act and, assuming that not all convicted prisoners are to be enfranchised, would involve difficult questions as to which side of the line disenfranchisement would fall".[139] He considered that this balancing act was a task for Parliament. On this basis he asserted:

> [T]hat any . . . relief which is intended to interfere with the process by which new legislation resulting from the consultation process is now put before and debated by Parliament, is inappropriate and is not to be granted.[140]

This conclusion stretches the rule protecting proceedings in Parliament to protect consultation documents produced by the government, and not by Parliament, from judicial comment simply on the basis that they may, in the future, form the basis of legislation, and stands in contrast to a long history of the judiciary commenting upon or agitating on behalf of reform proposals. For example, as the Constitutional Reform Bill was introduced into

130 See *Toner and Walsh* [2007] NIQB 18, para. 8.

131 Consultation Paper CP6/09, *Second Stage Consultation*, n. 5 above, p. 24.

132 Ibid. para. 9(iv).

133 See *Chester* [2009] EWHC 2923 (Admin), para. 48.

134 Sedley J, as he then was, once described the convention as providing for "a mutuality of respect between two constitutional sovereignties": *R* v *Parliamentary Commissioner for Standards ex parte Al Fayed* [1998] 1 WLR 669, p. 670. See also *Office of Government Commerce* v *Information Commissioner* [2008] EWHC 774 (Admin), para. 46, per Stanley Burnton J.

135 *R* v *HM Treasury ex parte Smedley* [1985] QB 657.

136 Ibid. p. 666.

137 Ibid. In deciding *ex parte Smedley* the Court of Appeal did proceed to consider whether the draft Order in Council in question (concerning an undertaking regarding payments to the then European Community) would be *ultra vires* the European Communities Act 1972.

138 See *Chester* [2009] EWHC 2923 (Admin), para. 48.

139 Ibid. para. 66.

140 Ibid. para. 52.

Parliament, senior judges actively criticised some of its proposals.[141] Even more strident judicial opposition was mounted to the ouster clause contained within the Asylum and Immigration (Treatment of Claimants, etc.) Bill 2003.[142] Although these examples involve extra-judicial activities,[143] Burton J's reading of the convention would prevent judges from commenting on the compatibility with the ECHR of any law reform proposals which might form the basis of future legislation.

Burton J buttressed this broad interpretation of the convention against interfering with proceedings before Parliament with a selective reading of decisions such as *Countryside Alliance*. In this case, Lord Bingham, rejecting the Countryside Alliance's attempt to challenge the ban on fox hunting, asserted that:

> The democratic process is liable to be subverted if, on a question of moral and political judgment, opponents of the Act achieve through the courts what they could not achieve in Parliament.[144]

Burton J accepted that, on the basis that prisoner disenfranchisement and fox hunting were both "matters of sensitive social policy",[145] an analogy could be drawn between these cases. He therefore justified his refusal to engage in a "democratic dialogue" regarding the reform process on the basis that it accorded with the duty of the courts not to subvert the democratic process. However, this analogy glosses over the fact that prisoner enfranchisement, in contrast to fox hunting, is not an issue which has been subject to the "recently expressed views of the democratically elected legislature".[146] The Representation of the People Act 1983 merely consolidated earlier legislation and later enactments such as the Representation of the People Act 2000 dealt with issues which were tangential to the voting rights of convicted prisoners.[147] Moreover, *Countryside Alliance* involved no clear infringement of any right, with Strasbourg ultimately rejecting claims that the ban on fox hunting infringed Article 8.[148] By contrast, the courts have consistently affirmed the constitutional importance of the right to vote in the United Kingdom.[149] In her speech in *Countryside Alliance*, Baroness Hale considered that, where an identified right has been infringed, the courts cannot abdicate their role in favour of action by Parliament:

> When we can make a good prediction of how Strasbourg would decide the matter, we cannot avoid doing so on the basis that it is a matter for Parliament. Strasbourg will be largely indifferent to which branch of government was responsible for the state of the domestic law.[150]

This rallying cry in favour of judicial action applies equally to the above discussion of the HRA's re-interpretive power. But if Burton J had reason to be wary of re-interpreting

141 See Lord Windlesham, "The Constitutional Reform Act 2005: the politics of constitutional reform" (2006) *PL* 35, p. 39.

142 See J McGarry, "Parliamentary sovereignty, judges and the Asylum and Immigration (Treatment of Claimants, etc.) Bill" (2005) 26 *Liverpool LR* 1, p. 9.

143 For an acknowledgment for the scope of advisory declarations as to the requirements of the law in judicial decisions, see *R (on the application of the Campaign for Nuclear Disarmament)* v *Prime Minister and others* [2002] EWHC 2777, para. 15, per Simon Brown LJ.

144 *R (on the application of the Countryside Alliance and others)* v *Attorney General* [2007] UKHL 52, para. 45.

145 *Chester* [2009] EWHC 2923 (Admin), para. 51.

146 *Countryside Alliance* [2007] UKHL 52, para. 124, per Baroness Hale.

147 See R Jago and J Marriot, "Citizenship or civic death? Extending the franchise to convicted prisoners" (2007) 5 *Web JCLI*.

148 See *Countryside Alliance* v *United Kingdom (2010)* 50 EHRR SE6, para. 46.

149 See *Watkins* v *Secretary of State for the Home Department* [2006] UKHL 17, para. 26, per Lord Bingham.

150 *Countryside Alliance* [2007] UKHL 52, para. 125.

s. 3 of the Representation of the People Act 1983 in *Chester*, he should at least have assessed whether the government's long-delayed reform proposals would have infringed the right to vote, particularly with regard to the proposals to continue the disenfranchisement of "post-tariff lifers", such as the claimant. Instead, Burton J withdrew from any putative constitutional dialogue. He chose, in the words of Tom Hickman, neither to "counteract protectively" against the infringement, nor to "interact productively" with the government's proposals.[151] In the HRA era, in which the United Kingdom's courts have a role alongside Parliament in addressing breaches of the ECHR, the courts need to develop the constitutional convention protecting parliamentary proceedings in a manner which permits judicial comment upon the executive's reform proposals before they come before Parliament.[152] Such a development would accord with the rule that in areas of "high constitutional importance"[153] the executive does not enjoy a wide discretionary area of judgment. As for Burton J's arguments against this course on the basis of the separation of powers, Parliament ultimately has the ability to depart from the recommendations of the courts regarding such reform proposals if it chooses to do so. In doing so, however, it would have to "squarely confront" the human rights concerns at issue.[154]

Conclusion: separation of powers or splendid isolation?

The prisoner enfranchisement cases exemplify the ambiguity which has resulted from the domestic courts' failure to develop "a clear and stable conception of the prisoner's legal status".[155] Governments have been able to manipulate the continuing uncertainty of many members of the judiciary regarding the legal status of prisoners, even in the HRA era. The domestic cases examined in this article, seem to bear out Karl Llewellyn's injunction not to underestimate the potential for first instance courts to tackle cases differently from the approach adopted by the highest court,[156] at least where s. 2 of the HRA is at issue. Even as appellate courts adopt ever more nuanced approaches to the issue of following Strasbourg jurisprudence, many first instance judges remain spellbound by the clarity and simplicity of the "mirror" principle expounded by Lord Bingham in *Ullah*.[157] Such reluctance to follow all but the most obviously "clear and constant"[158] Strasbourg jurisprudence harkens back to an earlier era of jurisprudence, when judges accepted that "decisions to be made as to the public interest are not such as courts are fitted or equipped to make".[159]

Furthermore, these cases emphasise the weaknesses in the system of remedies established by the HRA. Aileen Kavanagh rejects analyses of the operation of the HRA which identify "highly polarised" preferences for s. 3 or s. 4 amongst the judiciary. Instead, she asserts that judges and academics "will assess the appropriateness of the judicial choice between section 3 and section 4 in light of the facts and the context of the individual case".[160] Her conclusion, however, does not allow for the possibility that, despite the existence of a breach, the courts will use their discretion to exercise neither of these

151 Hickman, "Constitutional dialogue", n. 121 above.

152 Such judicial activity could replace, for matters within the scope of the ECHR, the moribund advisory role of the Judicial Committee of the Privy Council under the Judicial Committee Act 1833, s. 4.

153 *Kebilene* [2000] 2 AC 326, p. 381, per Lord Hope.

154 See *R* v *Secretary of State for the Home Department ex parte Simms* [1999] 3 All ER 400, p. 412, per Lord Steyn.

155 L Lazarus, "Conceptions of liberty deprivation" (2006) 69 *MLR* 738, p. 769.

156 K Llewellyn, *The Bramble Bush* (New York: Oceana 1951), p. 90.

157 *Ullah* [2004] UKHL 26.

158 *Alconbury* [2001] UKHL 23, para. 26.

159 *Gouriet* v *Union of Post Office Workers* [1978] AC 435, p. 482, per Lord Wilberforce.

160 Kavanagh, *Constitutional Review*, n. 107 above, pp. 123–4.

choices. The case of *Chester* evidences the potential powerlessness of the judiciary after a first declaration of incompatibility has been issued in instances where it is subsequently ignored by ministers on the basis of political expediency.[161]

The failure of the HRA's system of remedies in these cases was compounded by the reluctance of the courts to consider the limitations of the reform proposals in light of the Grand Chamber's advice in *Hirst*. This reluctance underlines the danger in overstating the degree to which the United Kingdom's constitutional law has entered an era of "interaction" and indeed "interdependence" between the three branches of government.[162] As Rodney Austin asserts, the courts are not a monolithic institution and "the new constitutionalism is not universally embraced by the judiciary".[163] Judges like Burton J appear happy to leave the role of assessing legislative proposals to Parliament's JCHR, with its "keen eyes open for any interference with human rights".[164] Indeed, in the many situations where ministers respond in a timely and appropriate manner to the JCHR's often "quasi-judicial" advice,[165] there is little need for judicial involvement in the process of law reform. Nonetheless, when an issue proves to be "legally straightforward, but politically difficult",[166] governments are adept at sidelining the criticisms of parliamentary committees. In such circumstances, where an issue is as important to democracy in the United Kingdom as prisoner enfranchisement, judicial decisions which remain aloof from the controversy draw less upon respect for the separation of powers, than upon a desire for splendid isolation from certain constitutional disputes.[167]

161 See T Campbell, "Incorporation through interpretation" in T Campbell, K Ewing and A Tomkins (eds), *Sceptical Essays on Human Rights* (Oxford: OUP 2001), pp. 79, 81–2.

162 Kavanagh, *Constitutional Review*, n. 107 above, p. 406.

163 R Austin, "The new constitutionalism, terrorism and torture" (2007) 60 *CLP* 79, p. 82.

164 *Chester* [2009] EWHC 2923 (Admin), para. 49.

165 F Klug and H Wildbore, "Breaking new ground: the Joint Committee on Human Rights and the role of Parliament in human rights compliance" (2007) *EHRLR* 231, p. 243.

166 JCHR, 31st Report, n. 48 above, para. 62. In such circumstances Francesca Klug and Helen Wildbore assert that "[i]t is not that the advice of the JCHR is held in less esteem than that of [other] scrutiny committees, but that Ministers are aware of the rather more discretionary and controversial nature – and values-base – of many human rights assessments": Klug and Wildbore, "Breaking new ground", n. 165 above, p. 240.

167 Subsequent to the completion of this article, the Court of Appeal heard Peter Chester's appeal from Burton J's decision (*R (Chester)* v *Secretary of State for Justice* [2010] EWCA Civ 1439). Again, the court refused to use its powers under the HRA to address prisoner disenfranchisement. It upheld the High Court's decision, with Laws LJ asserting, at para. 35, that reform of the voting ban on prisoners "is a political responsibility, and that is where it should remain".

NILQ 62(1): 75–98

Compensating defrauded shareholders in insolvency: is parity the answer?

BLANCA MAMUTSE*

University of Surrey

Abstract

This paper studies recent developments in Australian and US law permitting compensation for defrauded investors. For insolvent companies, these developments have drawn attention to the possibility of investor claims being satisfied on parity with the claims of ordinary unsecured creditors. This paper proposes that such a shift may be justified on the basis of a modern perspective of the principles underpinning corporate law. However, account must also be taken of the more practical implications which may hinder a widespread acceptance of parity.

Introduction

A notable development in the field of investor protection law over the past few years has been the extension of the concept of rateable (or *pari passu*) distribution within the class of ordinary unsecured creditors to accommodate the interests of defrauded shareholders.[1] This is due to the phenomenon of shareholder claims, defined as claims for damages against a company by a subscriber for or purchaser of its shares, where the claimant relies upon misleading or deceptive conduct of the company, or other wrongful acts or omissions on its part as being causative of his or her loss.[2] At first glance, it would seem that such an approach is contrary to the rule that debts owed to a member of a company in his or her capacity as such (*qua* member) may not be satisfied before the claims of other creditors have been met.[3] This enforces the principle that:

> . . . the rights of members *as members* come last, i.e. rights founded on the statutory contract are, as the price of limited liability, subordinated to the rights

* Lecturer in Law, University of Surrey, UK. I am grateful to my colleague Alison Clarke for her insightful comments on an earlier draft of this paper. A condensed version of this paper was presented at the Society of Legal Scholars' annual conference in September 2011: I am indebted to the participants of the subject session for their helpful suggestions. I take responsibility for any errors.

1 See landmark decision of Australia's highest court in *Sons of Gwalia Ltd* v *Margaretic* [2007] HCA 1; and changes brought about in US by Sarbanes–Oxley Act – both discussed below. The *Gwalia* judgment has led Australian Federal Government to introduce new legislation, also discussed below.

2 *Sons of Gwalia* [2007] HCA 1, para. 34.

3 See, for example, s. 74(2)(f) of the UK Insolvency Act 1986. See, however, s. 655 Companies Act 2006 – a shareholder's claim against the company for damages or other compensation is not barred by mere fact of his or her ownership of (or entitlement to) shares in the company.

> of creditors based on other legal causes of action. The rationale . . . is to ensure that the rights of members as such do not compete with the rights of the general body of creditors.[4]

In English law, this signals a distinction which is made in upholding shareholder claims, between sums claimed *qua* member arising from the statutory contract, and claims in respect of which membership is an essential qualification for acquiring the claim but not the foundation of the cause of action.[5] On the other hand, the approaches recently adopted in the US and in Australia drive us to review the justifications for the established rule in the light of modern perspectives of corporate relationships. It is a matter of interest to lawyers in other jurisdictions, as the financial downturn has resulted in claims for securities loss being pursued against companies in foreign courts, particularly in the US.[6] These developments also provide an important insight into the intersection between the insolvency distribution rules and market regulation. These underlying questions, and the connection between them, are helpfully condensed in this extract from the decision of the High Court of Australia in the matter of *Sons of Gwalia Ltd* v *Margaretic*:

> [M]odern legislation . . . has extended greatly the scope for "shareholder claims" against corporations, with consequences for ordinary creditors who may find themselves, in an insolvency, proving in competition with members now armed with statutory rights. Corporate regulation has become more intensive, and legislatures have imposed on companies and their officers obligations, breach of which may sound in damages, for the protection of members of the public who deal in shares and other securities. This raises issues of legislative policy. On the one hand, extending the range of claims by shareholders is likely to be at the expense of ordinary creditors. The spectre of insolvency stands behind corporate regulation. Legislation that confers rights of damages upon shareholders necessarily increases the number of potential creditors in a winding up. Such an increase normally will be at the expense of those who previously would have shared in the available assets. On the other hand, since the need for protection of investors often arises only in the event of insolvency, such protection may be illusory if the claims of those who are given the apparent benefit of the protection are subordinated to the claims of ordinary creditors.[7]

The policy challenge described in this statement becomes more apparent when considered against the backdrop of the established justifications for subordinating shareholder claims in insolvency.

---

4 *Soden* v *British & Commonwealth Holdings plc* [1998] AC 298, per Lord Browne-Wilkinson, para. 324 (original emphasis). The statutory contract referred to is now set out in s. 33(1) of the UK Companies Act 2006 (formerly s. 14(1) of the Companies Act 1985): "The provisions of a company's constitution bind the company and its members to the same extent as if there were covenants on the part of the company and of each member to observe those provisions."

5 See *Soden* v *British & Commonwealth Holdings plc* [1998] AC 298, based on a construction of s. 74(2)(f) Insolvency Act 1986.

6 See, for instance, M Herman, "Pensioners hire Cherie Blair to sue Sir Fred Goodwin and RBS", *The Times*, 16 March 2009; C Seib and M Waller, "Goldman Sachs case opens lawsuit floodgates on Wall Street", *The Times*, 18 April 2010. However, the US Supreme Court's recent decision in *Morrison* v *National Australia Bank Ltd* 2010 US LEXIS 5257 denying jurisdiction on the part of US courts to determine "f-cubed" cases (involving claims brought by foreign investors against foreign defendants in connection with transactions on a foreign exchange) may strongly limit the number of claims with a foreign element in future.

7 [2007] HCA 1, para. 18, per Gleeson CJ.

The basis for subordinating shareholder claims in insolvency

Historically, the subordination of shareholder claims in insolvency arose from the necessity to adapt laws to the creditors of corporations as distinct from the creditors of individual bankrupts.[8] This concern was reflected in two principles: firstly, that corporate debts should be paid before distributions are made amongst shareholders – a rule which has its origins in the conception of corporate assets as a trust fund for creditors.[9] The second was the notion that a shareholder is precluded from denying his or her liability to contribute to the assets of an insolvent company on the ground that the shareholder was induced to buy the shares by fraud.[10] In modern insolvency law shareholders are referred to as the "cushion on which all other creditors rest",[11] and, accordingly, equity claims in general are subordinated to the payment of other creditors. The justifications for subordination may be seen to be founded on the perception of this cushion, and can broadly be categorised as being based on contract, creditor reliance and equity. Contractually, the relationship between the shareholder and the company entails that the shareholder as a member undertakes to contribute a certain amount in satisfaction of its debts and liabilities, and it is arguably inconsistent with this position to permit the shareholder to lay claim to any part of those assets while external creditors remain unpaid.[12] However, where a contract to purchase shares is induced by fraud, it is voidable[13] and shareholders are not barred from pursuing their remedies against the company by reason of their membership.[14] The reliance aspect expresses the need, in ascribing a lower priority to the shareholder's contractual claims or damages award, to take account of the relative positions of shareholders and creditors. In particular, whereas shareholders enjoy the benefits of participation in the success or prosperity of an enterprise, creditors extend credit to an entity in (at least partial) reliance upon their perception of the equity "cushion" provided by the shareholders' investments, and would be prejudiced by a dilution of the capital reserves available to repay them if these were to be applied equally to the payment of shareholders.[15] In an insolvency situation, where there clearly remains no prospect of recovery from the equity cushion, the creditors shift the focus of their reliance to the expectation of priority over equity claims when the debtor's estate is divided up.[16] From the perspective of achieving equity, subordination furthermore ensures that innocent creditors do not bear the economic burden of shareholder fraud remedies against the debtor company, which was involved in the

8 E S Hunt, "The trust fund theory and some substitutes for it" (1902) 12 *Yale LJ* 63.

9 *Wood* v *Dummer* 3 Mason 309. Hunt, "The trust fund", n. 8 above, discusses whether the trust applies to a solvent company, or if it is limited to insolvent bodies (pp. 73–4).

10 *Oakes* v *Turquand* [1861–73] All ER Rep 738.

11 P Wood, *Law and Practice of International Finance* (London: Sweet & Maxwell 2008), pp. 5–19.

12 *Houldsworth* v *City of Glasgow Bank* [1874–80] All ER 333, at 335; *In re Addlestone Linoleum Co* (1887) 37 Ch D 191 at 206; R Goode, *Principles of Corporate Insolvency* 3rd edn (London: Sweet & Maxwell, 2005), p. 199.

13 *Oakes* v *Turquand* [1861–73] All ER Rep 738.

14 See, for instance, UK Companies Act 2006, s. 655: "A person is not debarred from obtaining damages or other compensation from a company by reason only of his holding or having held shares in the company or any right to apply or subscribe for shares or to be included in the company's register of members in respect of shares."

15 Observed by Ebel J in *Re Geneva Steel Co* (2002) 281 F3d 1173 at 1176 and 1179, a judgment in which the history of §510(b) of the US Bankruptcy Code is comprehensively discussed (Title 11, Chapter 11, USC). As will be seen below, this provision expressly subordinates claims for rescission or damages arising from the purchase or sale of shares, to payment of secured and unsecured creditors. See also M E Sprouse, "A collision of fairness: Sarbanes–Oxley and s. 510(b) of the Bankruptcy Code" (2005) 24–8 *American Bankruptcy Institute Journal* 8. See also K B Davis, "The status of defrauded securityholders in corporate bankruptcy" (1983) 1 *Duke Law Journal* 1, pp. 11–12.

16 Davis, "The status", n. 15 above, pp. 19–22.

wrongdoing and may have benefited from it.[17] A party who has irrevocably adopted the liabilities of a shareholder cannot appear to claim out of the debtor company's assets a sum not included in the debts and liabilities, to the payment of which he or she as a shareholder, had agreed that those assets should be devoted.[18] "Equity" may also be said to operate in the sense of recognising the extinction of shareholders' privileged status on the insolvency of a company:

> Shareholders' ample and superior statutory rights, their voluntary abdication of control over their investment in favour of their appointees, the directors, who have large statutory and constitutional discretions and obligations in the application of it, their rights of intervention, their rights to proceed against the directors personally as well as the company in some circumstances, their statutorily mandated limited liability, especially that, and their rights to participate in the bounty of any successes, sit uncomfortably with the notion that s 563A gives them equal billing, on the failure of the company, with ordinary creditors.[19]

Thus, an analysis of the relationship between shareholders and the company, between the creditors and the company, and between creditors and shareholders leads to the conclusion that subordination of shareholder claims is justified on contractual and equitable grounds. This is so although insolvency law in general[20] and particularly the principle of pro rata distribution[21] do not – as a broad rule – allow claims to be favoured on the basis of factors such as their origin or relative merits, the nature of the claimant or its relationship with the debtor. With respect to claims for loss resulting from the sale or purchase of the debtor company's shares, the question arises whether a merger should be allowed between the interests of unsecured creditors as debt-holders and the interests of shareholders as contractual or tort claimants. Or, on the other hand, if a separation should be maintained on the basis of their respective identities and the terms on which they are deemed to have contracted with the corporate debtor. The latter approach allows some consistency to be achieved in treating shareholders as shareholders regardless of whether their entitlement in a particular case is analogous to that of creditors, and continuing to accord priority to the creditors. The experience of the US and Australia, outlined below, shows how investor interests may encroach on the distribution rights of creditors, either as a result of legislative developments or by means of judicial construction.

Mandatory subordination in the United States and the impact of the Sarbanes–Oxley Fair Funds requirement

In the United States Bankruptcy Code, the main principle governing distribution in Chapter 7 liquidations as well as Chapter 11 reorganisations is that of "absolute priority", whereby claims that are classed as being of a higher priority are entitled to payment in full before the lower priority claims are met.[22] The Chapter 7 and Chapter 11 procedures are

17 Davis, "The status", n. 15 above, at p. 16.

18 *In re Addlestone Linoleum Co* (1887) 37 Ch D 191, at 205.

19 Callinan J's dissenting judgment in *Sons of Gwalia Ltd* v *Margaretic*, n. 1 above, para. 242, referring to s. 563A of the Australian Corporations Act 2001: "Payment of a debt owed by a company to a person in the person's capacity as a member of the company, whether by way of dividends, profits or otherwise, is to be postponed until all debts owed to, or claims made by, persons otherwise than as members of the company have been satisfied."

20 Davis, "The status", n. 15 above, pp. 16–18.

21 Also known as the *pari passu* principle – see M Bridge, "Collectivity, management of estates and the *pari passu* principle in winding-up" in J Armour and H Bennett (eds), *Vulnerable Transactions in Corporate Insolvency* (Oxford: Hart 2003), ch. 1, para. 1.2.

22 Title 11, Chapter 7, USC §726; and Title 11, Chapter 11, USC §1129(b)(2)(B)(ii).

furthermore connected by the inclusion, among the requirements for a court to confirm a Chapter 11 plan, of a stipulation that claimants shall receive under the plan an amount no less than they would receive in a Chapter 7 liquidation.[23] Consequently, the enforcement of the priority structure set out in Chapter 7 is essential for the proper confirmation of a Chapter 11 reorganisation plan.[24]

In terms of this priority structure, the claims of unsecured creditors are to be satisfied before those of shareholders. Among the ordinary (non-preferential) unsecured creditors, payment of their claims is to be made on a pro rata basis.[25] Thus, *pari passu* treatment operates within the framework of absolute priority, and is acknowledged as the "equality of distribution principle".[26] Of particular interest is §510(b) of the Code which provides that claims for the rescission of a purchase or sale of a security of the debtor company, or damages arising from the purchase or sale thereof, or claims for reimbursement or contribution for a sale or purchase, are subordinated to the payment of secured and unsecured claims. By this rule of mandatory subordination, shareholders are precluded from recovering from the company in respect of claims arising in contract or tort on parity with unsecured creditors.[27] The possibility of this form of shareholder elevation is clearly excluded, and the equality of distribution principle continues to apply as between the general creditors.

With the introduction of the American Competitiveness and Corporate Accountability Act (referred to herein as the Sarbanes–Oxley Act),[28] in particular its Fair Funds provision §308(a), this distribution regime has undergone a modification, as the Securities and Exchange Commission (SEC) is entitled to direct that civil penalties recovered in respect of securities violations be added to a disgorgement fund for the benefit of victims of such violations. The primary purpose of disgorgement is to deter violations of the securities laws by depriving violators of their illegal profits.[29] Thus, although disgorged funds may often go to compensate securities fraud victims for their losses, such compensation is a distinctly secondary goal, and the measure of disgorgement need not be tied to the losses suffered by defrauded investors.[30] Civil penalties further the deterrent role of disgorgement by allowing the SEC to impose a financial penalty which may amount to the gross monetary gain from the securities fraud.[31] Consequently, neither mechanism entails the enforcement or collection of a debt as such.

A change has come about in that, while the SEC formerly had only discretion to distribute recoveries from disgorgement to injured investors and was obliged to transmit civil penalties to the US Treasury,[32] the Fair Funds provision now permits it to add the civil

23 Title 11, Chapter 11, USC §1129(A)(7)(ii).

24 D A Henry, "Subordinating subordination: WorldCom and the effect of Sarbanes–Oxley's Fair Funds provision on distributions in bankruptcy" (2004) 21 *Bankruptcy Developments Journal* 259, p. 270.

25 Title 11, Chapter 7, USC §726(b).

26 Z Christensen, "The Fair Funds for Investors provision of Sarbanes–Oxley: is it unfair to the creditors of a bankrupt debtor?" (2005) *University of Illinois Law Review* 339, p. 347.

27 Henry, "Subordinating subordination", n. 24 above, p. 272.

28 2002 Pub L No 107–204, 116 Stat. 745.

29 SEC v *Fischbach Corporation*, 1333 F3d 170, at 175. See also *The Investor's Advocate: How the SEC protects investors, maintains market integrity, and facilitates capital formation*, United States SEC, available at www.sec.gov/about/whatwedo.shtml.

30 *SEC* v *Fischbach Corporation*, 1333 F3d 170, at 175–6.

31 Securities Enforcement Remedies Act and Penny Stock Reform Act of 1990, 15 USC §78u(d)(3).

32 See *Official Committee of Unsecured Creditors of WorldCom Inc.* v *SEC* (2006) 467 F3d 73, at 81–2, and 15 USC §77t(d)(3)(A).

penalties proceeds to the disgorged funds and distribute these monies to the victims of the fraud. It is recognised that this has made "another potentially large category of funds available for compensation".[33] This enhanced power of the SEC supports the intention underlying the Act, to "protect investors by improving the accuracy and reliability of corporate disclosures made pursuant to the securities laws"[34] and "crack down on fraud and wrongdoing".[35] Although the operation of the provision is not limited to companies that are insolvent, the application of the Fair Funds term in a bankruptcy has meant that a shareholder whose claim against the debtor company falls for mandatory subordination in terms of §510(b) may be eligible to receive compensation as a victim of fraud *pari passu* with unsecured creditors.[36] In other words, the SEC's share in the bankruptcy estate as a general unsecured creditor may be distributed to the defrauded shareholders rather than being paid over to the US Treasury.[37] In this way, the absolute priority rule is bypassed and, to the extent that the distinguishable (and in some respects opposing) interests of shareholders and unsecured creditors are unusually aligned,[38] its implications for *pari passu* treatment merit consideration.

In terms of general insolvency principles, this is clearly a departure from the established framework[39] whereby the payment of equity debts is made after distribution to unsecured creditors. In light of the specific requirement of mandatory subordination of shareholder claims under the US Bankruptcy Code, whereby even claims going to the root of the shareholders' relationship with the debtor company[40] and of a potentially tortious nature are categorised as equity claims, this development is even more striking. Furthermore, in the field of corporate rescue law, as compared with the traditional focus of a reorganisation process, such as Chapter 11, on the debtor's financial circumstances and recovery prospects,[41] Fair Funds uniquely diverts attention towards the obligation to take account of the needs of a particular group of creditors.

The consequences of implementing the Fair Funds provision §308(a) in bankruptcy were recognised by Rakoff J in assessing the penalty of the SEC in its case against WorldCom Inc.[42] He affirmed that, under the bankruptcy laws, the SEC's penalty claim was

33 V Winship, "Fair Funds and the SEC's compensation of injured investors" (2008) 60 *Florida Law Review* 1103, p. 1110. She argues, moreover, that since the passage of the Sarbanes–Oxley Act, penalties serve the dual purpose of deterring securities law violations and compensating harmed investors (p. 1118).

34 Preamble to the Sarbanes–Oxley Act.

35 Statement of President George W Bush on House of Representatives Action on the proposed "Sarbanes–Oxley Act of 2002", 25 July 2002.

36 Christensen, "Fair Funds", n. 26 above, illustrates this by example of a tort claim by the purchaser of an equity security that arises from the debtor's fraudulent activity. In the absence of a Fair Funds requirement and by virtue of mandatory subordination, this claim would not achieve the same status as a general unsecured claim even though in principle a holder of a tort claim would obtain the status of a general unsecured creditor: p. 348.

37 Ibid. pp. 353–4.

38 Sprouse, "A collision", n. 15 above, notes that this distribution may be received by the stockholder claimant "to the direct detriment of unsecured creditors and in potential contravention of provisions of the Code, including §510(b) and the 'absolute priority rule'": p. 8.

39 See, generally, Wood, *Law and Practice*, n. 11 above, at 5-02, and in particular Title 11, Chapter 7, USC §726(a), which sets out the order of priority for claims, subject to the requirement of mandatory subordination according to §510.

40 For instance, whether the purchase of shares is tainted by misrepresentation.

41 A Keay and P Walton, *Insolvency Law: Corporate and personal* 2nd edn (Bristol: Jordans 2008), para. 1.3.

42 *SEC* v *WorldCom Inc.*, 273 F Supp 2d 431, United States District Court for the Southern District of New York, affirmed on appeal: *Official Committee of Unsecured Creditors of WorldCom Inc.* v *SEC* 467 F3d 73 (2006) (US Court of Appeals for 2nd Circuit).

treated simply as another claim by one of many unsecured creditors.[43] However, he acknowledged that while §308(a) gave the SEC an opportunity to pay any penalty it recovered to shareholder victims rather than to the US Treasury, it could not properly premise the size of its penalty on the basis of such disbursement. This move would arguably run afoul of the mandatory subordination provisions of the Code, given that there was no suggestion in §308(a) that Congress intended to accord a greater priority to shareholders in bankruptcy than they had previously enjoyed. Rakoff J, however, considered it to be acceptable for the SEC to give its penalty recovery to shareholder victims or take *some* account of shareholder loss in formulating the size and nature of its penalty.[44] The approved settlement provided that the penalty recoveries would be directed to defrauded shareholders in accordance with the Fair Funds provision.

Rakoff J's decision was upheld on appeal following objections by an official committee of WorldCom's unsecured creditors (the Committee) to the exercise of the district court's discretion in approving the plan.[45] The Committee took issue with the SEC's distribution plan on the ground that, in the absence of sufficient funds to compensate all the victims of WorldCom's fraud, several groups of investors were excluded from sharing in the amount collected from WorldCom. Among the investors excluded were those who had recovered 36 per cent or more on their claims under the Chapter 11 reorganisation plan or through the sale of their securities, and investors who made a net profit on their combined purchase and sales of WorldCom securities during the period in which the fraud occurred. The Committee further contended that the ordinary judicial test for reviewing the distribution of disgorged profits – that is whether the distribution was conducted in a "fair and reasonable" manner[46] – should not apply to the review of Fair Funds plans. The Committee especially objected to the exclusion of creditors who had received more than 36 per cent in the bankruptcy proceedings or through the sale of their WorldCom securities as flying "in the face of the strong public policy that puts the rights of bondholders ahead of those of shareholders".[47] With regard to the implications for the relationship between creditors, the appeal Court recognised, in the same manner as the district court, the tension between the priority assigned to claims under the Bankruptcy Code and the Fair Funds provision, which empowered the SEC to distribute funds among injured investors outside the bankruptcy proceeding.[48] The court saw no indication in the Fair Funds provision, however, that the SEC must follow the Bankruptcy Code's claim priorities when developing a distribution plan, and it considered that in the absence of such an indication it was not its role to mitigate this tension.[49] It found that the district court was required only to determine that the SEC's distribution plan fairly and reasonably distributed limited Fair Fund proceeds among potential claimants, and it had properly exercised its discretion in approving the plan.[50] In this way, a distribution outcome in the context of bankruptcy which did not accord with the equality of distribution principle (or *pari passu* treatment) was rendered

43 *Official Committee of Unsecured Creditors of WorldCom Inc.* v *SEC* 467 F3d 73 (2006) (US Court of Appeals for 2nd Circuit), p. 434.

44 Ibid. emphasis added. In the result, the penalty proposed and approved by the court was $750 million, 75 times greater than any prior such penalty.

45 Ibid. The question whether the committee had exceeded its statutory authority as a creditors' committee in bringing the proceedings was argued before the court, pp. 77–81. Provision governing powers of creditors' committees: Title 11, Chapter 11, USC §1103.

46 The test applied in the court *a quo*: *SEC* v *WorldCom Inc.*, n. 42 above.

47 *Official Committee*, n. 42 above, p. 85.

48 Ibid.

49 Ibid. p. 85.

50 Ibid.

acceptable by framing the issue for resolution not as whether the SEC was constrained to observe the Bankruptcy Code's priorities in making distributions; but rather by inquiring whether it fell outside the scope of the SEC's authority to make distributions that were inconsistent with such priorities.

The actual compromise and settlement between WorldCom and the SEC was approved by the Bankruptcy Court for the Southern District of New York.[51] Gonzalez J refrained from deciding whether the ultimate distribution to securities holders contemplated by the settlement violated the absolute priority rule and mandatory subordination under the Code. In his opinion, even if this were found to be the correct legal interpretation from a bankruptcy standpoint, there remained a number of legal issues to be addressed. The nature of these issues, as well as any other issues that may be raised in litigation to subordinate the claim, was such as to furnish sufficient doubt as to the outcome of any litigation to subordinate the claim of the securities holders. He took this uncertainty among other factors as providing support for the settlement, and went further to state that in considering the approval of a settlement the court was not required to resolve the "underlying legal issues" related to the settlement.[52] In the court's determination, the settlement fell within the range of reasonableness and was fair and equitable and in the best interests of the debtors' estates.[53] As noted above, the mere fact of inconsistency with the absolute priority rule did not influence the court against approving the settlement, and this incompatibility did not have a bearing on the "fairness and reasonableness" of the settlement that had been reached – it was acknowledged as constituting no more than a legal issue and was thus separable from the matters to be taken account of in sanctioning the settlement.

This reluctance to resolve the contradiction between approving a settlement with the SEC, whereby a victims' restitution fund would be established from the settlement proceeds, and the likelihood that the distribution of the same monies would potentially allow a division contrary to §510(b) and *pari passu* distribution among ordinary unsecured creditors was echoed in *Re Adelphia Communications*.[54] Gerber J acknowledged that the victims' restitution fund would be distributed largely to victims who were equity security holders or investors of debt securities who would find themselves subordinated under §510(b) if they asserted claims in Chapter 11 cases. But he did not consider this to be a satisfactory basis for disapproving the settlement, as, firstly, these parties would not be sharing in the assets of the estate but under a plan created and owned by the government; and, secondly, the uncertainty surrounding the disbursement of Fair Funds vis-à-vis the absolute priority rule could be taken as supporting the view that settlement would be appropriate in the circumstances.[55] It is strongly indicative of the very real difficulty that the courts have faced in these matters that as with Gonzalez J's judgment in the *WorldCom* matter, Gerber J took a negative element – namely the uncertainty arising from implementing a Fair Funds distribution in the light of the absolute priority structure – as evidence of the desirability of a settlement.

It may be seen that the legislature's omission to reconcile the effect of the Fair Funds requirement with the current bankruptcy structure has been matched by the judicial demurral highlighted above. The imperative of facilitating a settlement in the cases outlined above has enabled the courts to manoeuvre around the difficulty of making a choice between the competing objectives of eliminating fraud and preserving the established

51 *In re WorldCom Inc.*, Case No 02 B 13533 (AJG): August 6, 2003.

52 Ibid. p. 4 of judgment.

53 Ibid.

54 *In re Adelphia Communications Corporation* (2005) 327 BR 143.

55 Ibid. pp. 168–71.

bankruptcy distribution scheme. Consequently, no views have emerged as to which is to be upheld or promoted over the other, and this underscores the uncertainty attendant on the conception of *pari passu* treatment as a matter of public policy:[56] it does not anticipate situations where having regard to the public interest might lead to the conclusion that the enforcement of equal treatment should be extended to non-creditor parties. The apparent inconsistency between §510(b) of the Code and the Fair Funds provision has been widely noted,[57] and in the absence of any statutory indication as to whether the application of the Fair Funds provision should alter the established distributional priority scheme in the Bankruptcy Code,[58] commentators have advanced some persuasive arguments regarding whether these dissonances should be viewed in a positive or negative light. Before these arguments may be considered in greater detail, it is necessary to introduce the developments in Australian law which temporarily produced a similar outcome.

Sons of Gwalia Ltd v *Margaretic* in Australia: the High Court's decision and its consequences

Sons of Gwalia Ltd v *Margaretic* is a recent decision of Australia's highest court,[59] the salient facts of which are as follows. In August 2004, Mr Margaretic (M) bought 20,000 fully paid ordinary shares in the capital of Sons of Gwalia Ltd (Gwalia). The company was listed on the Australia Stock Exchange and it was from this market that M made his purchase. A few days after M was entered on the register of the company's members, administrators were appointed to distribute the assets of the company as in a winding-up.[60] Upon the appointment of the administrators, the Gwalia shares bought by M became completely worthless. M claimed compensation from Gwalia, alleging a breach of the securities laws. The administrators of Gwalia sought to prevent him from proving his claim in the deed of company arrangement as a creditor, rather than a member. Under s. 563A of the Australia Corporations Act 2001:

> Payment of a debt owed by a company to a person in the person's capacity as a member of the company, whether by way of dividends, profits or otherwise, is to be postponed until all debts owed to, or claims made by, persons otherwise than as members of the company have been satisfied.[61]

---

56 A connection made in the House of Lords' judgment in *British Eagle International Airlines* v *Compagnie Nationale Air France* [1975] 1 WLR 758 (HL) in refusing to uphold an airline netting arrangement on British Eagle's insolvency. See, however, more recent authorities declining to give effect to an independent (i.e. non-statutory) notion of public policy in relation to the rule: *Perpetual Trustee Co. Ltd* v *BNY Corporate Trustee Services Ltd* [2009] EWCA Civ 1160 (England and Wales Court of Appeal); *International Air Transport Association* v *Ansett Australia Holdings Ltd* [2008] HCA 3 (High Court of Australia).

57 Christensen, "Fair Funds", n. 26 above; Sprouse, "A collision", n. 15 above; and Henry, "Subordinating subordination", n. 24 above. For a discussion from an international perspective, see J Sarra, "From subordination to parity: an international comparison of equity securities law claims in insolvency proceedings" (2007) 16 *International Insolvency Review* 181.

58 Christensen, "Fair Funds", n. 26 above, pp. 369–75.

59 *Sons of Gwalia Ltd* v *Margaretic* [2007] HCA 1. Due to the constraints of space, the procedural history of the litigation is not set out in this paper. For a helpful summary, combined with a detailed analysis of the High Court judgment, see A Hargovan and J Harris, "The shifting balance of shareholders' interests in insolvency: evolution or revolution?" (2007) 31 *Melbourne University Law Review* 591.

60 Under s. 463A(1) of the Corporations Act 2001.

61 Although the matter was argued on the basis that the issue for decision turned on the proper construction of s. 563A, this was reinforced by cl. 4(2)(d) of the deed of company arrangement entered into by Gwalia: "For the avoidance of doubt, payment of any debts or liabilities owed by the Company to Members in the Members' capacity as a member of the Company, whether by way of dividends, profits or otherwise are, to the extent contemplated by Section 563A of the [Corporations Act 2001] and the general law, to be postponed until all debts owed to, or claims made by, Creditors have been satisfied."

The questions for determination were therefore whether M's claim was a provable debt, and if so whether it ranked for payment with non-shareholder creditors, or was postponed to the satisfaction of their claims. Its ranking with non-shareholder creditors would depend on whether M's claim fell within the s. 563A subordination of debts owed to a member as a member. Thus, although there is no statutory provision equivalent to §510(b) of the US Bankruptcy Code prescribing mandatory subordination of the sort of claim brought by M, attention focused on the nature of the debts which could be considered as owing to a person *qua* member under s. 563A.

Two important factors were noted that had a bearing on this case. The first was the historical significance of s. 563A, which could be traced back to a time when the separate legal personality of a company had not been fully recognised, and the distinction between corporations and partnerships was less marked. It was an established rule of partnership law that a partner in a bankrupt firm could not prove in competition with debts of outside creditors upon a dissolution, as this would permit him or her to diminish partnership assets to the prejudice of the firm's creditors, who were also his or her own creditors. This rule now formed part of the conception of a company's existence as an entity separate from its members.[62] Another influential consideration, relevant to the raising and maintenance of capital, was the established principle that the creditors of a company which is being wound up have a right to look to the paid-up capital as the fund out of which their debts are to be discharged.[63] This common law principle had its origins in the nineteenth century,[64] and had been given statutory effect in successive Companies Acts. These factors provided the background to the operation of s. 563A and demonstrated the wider implications of the decision, beyond the resolution of this matter.

The majority of the High Court concluded that M's claim was not a debt owed to him in his "capacity as a member" of Gwalia, whether by way of dividends, profits or otherwise. Accordingly, the claim was not to be postponed by s. 563A of the Corporations Act 2001 to claims made by "persons otherwise than as members of the company". In determining whether the type of claim brought by M fell to be subordinated under s. 563A, analysis centred on the interpretation of the phrase "in the person's capacity as a member of the company". The provision itself did not manifest any clear legislative policy, as compared with the mandatory subordination provision §510(b) of the US Bankruptcy Code. However, it was clear to the court that it did not embody a general policy that "members come last" in an insolvency:

> On the contrary, by distinguishing between debts owed to a member in the capacity as a member and debts owed to a member otherwise than in such a capacity [s. 563A] rejects such a general policy.[65]

The criteria for subordination were thus based on the character of the debt rather than the identity of the claimant. In this case, the claim made by M was not founded on any rights he obtained or obligations incurred by virtue of his membership of Gwalia. M did not seek to recover any paid-up capital or to avoid any liability to contribute to the company's capital. His membership of the company was not definitive of the capacity in which he made his

62 *Sons of Gwalia Ltd* v *Margaretic* [2007] HCA 1, paras 3–4.

63 Gleeson CJ expressed doubts regarding the modern significance of this notion: "Statutory manifestations of that principle have been modified over the years, and it may be doubted that it reflects the reality of modern commercial conditions, where assets and liabilities usually are more significant for creditors than paid-up capital. As Lord Browne-Wilkinson said in *Soden* v *British & Commonwealth Holdings plc* [1998] 2 LRC 225, at 232, it is "wholly irrelevant to the position of a member who has acquired fully paid shares on the market": para. 5.

64 *Trevor* v *Whitworth* (1887) 12 App Cas 409.

65 *Sons of Gwalia Ltd* v *Margaretic* [2007] HCA 1, para. 19. See also para. 119.

claim, and the obligation which he sought to enforce was not an obligation which the Corporations Act created in favour of a company's members. The obligations arose by virtue of Gwalia's conduct in relation to the statutory duties alleged to have been breached – in particular the prohibition against engaging in misleading or deceptive conduct – rather than from a legislative prescription of the rights and duties of company members. Gwalia's contravention of "statutory norms of behaviour"[66] rendered it liable to provide damages or other relief at the suit of any person who had suffered, or was likely to suffer, loss and damage as a result of the contravention. Accordingly, s. 563A did not apply to M's claim, as it was not a debt owed to him in his capacity as a member of Gwalia.

Callinan J dissented from this construction of s. 563A. He considered that the scope, objects and history of the Corporations Act 2001, the language of s. 563A and the context in which the provision appeared, the relevant case law and the desirability of maintaining coherence and fairness within the law, all pointed towards the construction that s. 563A precluded the treatment of M as a creditor on parity with other unsecured creditors. He drew attention to the conformity of this result with the typical shareholder/creditor relationship:

> . . . up to the point of insolvency, liquidation or administration of a company, its members enjoy superior opportunities, rights and advantages to creditors, yet the latter are no less likely to be disadvantaged by deceptive conduct of a company lying in a failure to comply with the continuous disclosure rules. There can be no doubt that the financial capacity of a company to satisfy its obligations to all of those who deal with or rely on it, is a matter of continuing interest and concern to them.[67]

He noted that parity would also distort inter-shareholder relationships, given that all shareholders of Gwalia had been equally wronged and induced by the company to hold on to their shares. Recent purchasers such as M would gain a large advantage over other, equally wronged, longer-term members if their claims were accorded the status of non-membership debts. He thus concluded that M's claim should be postponed to the satisfaction of all non-member creditor debts.

It may be seen how defrauded shareholders in the US and Australia respectively found themselves in similar positions; as a result of the Fair Funds legislation in one jurisdiction and of judicial construction in the other. Concern arose in Australia over the implications of the *Sons of Gwalia* decision, which included its potential to facilitate the participation of equity holders in insolvencies and workouts; encourage litigation by shareholders or the likely use of compensation claims as leverage in insolvencies; and an increase in the tendency of banks to request security when lending to a listed company.[68] Notwithstanding the recommendation by the national Corporations and Markets Advisory Committee not to reverse the effects of the *Sons of Gwalia* judgment,[69] the Australian government has released draft legislation overturning the decision. The Corporations Amendment (No 2) Bill 2010 explicitly provides that claims in relation to the buying, selling, holding or otherwise dealing in shares are to be ranked equally and paid after satisfaction of all creditors' claims. The consequences of the *Sons of Gwalia* decision with respect to reduced access to debt finance and the increased cost

66 *Sons of Gwalia Ltd* v *Margaretic* [2007] HCA 1, para. 10.

67 Ibid. para. 241. See also text accompanying n. 19 above.

68 See D Loxton, "Shareholder claims in corporate failure: why is Australia ditching *Gwalia*?" (2010) *Journal of International Banking and Financial Law* 244.

69 Corporations and Markets Advisory Committee, *Shareholder Claims against Insolvent Companies: Implications of the* Sons of Gwalia *decision* (Government of Australia, 2008). Hereafter referred to as the CAMAC Report.

and complexity of insolvent administrations, have thus been addressed.[70] To this extent, the debate in Australia regarding the use of the *qua*-member construction approach to confer parity on investors appears to be closed. As other jurisdictions take this direct, legislative approach to clarifying the position, certainty may be expected to dominate this area of law. However, this does not exclude the potential for mandatory subordination to be apparently undermined by another legislative instrument which is underpinned by similarly compelling policy reasons, as borne out by the Sarbanes–Oxley Fair Funds experience. For this reason, the consideration which follows of the arguments supporting parity focuses principally on the US position, although it incorporates more general observations which may be relevant to the legal position in other countries.

Can parity be justified as a progressive move in insolvency distribution?

The Fair Funds provision has been recognised as a positive development on the ground that distributing the funds of a bankrupt estate to shareholders as tort claimants is essential to a modern system of corporate governance.[71] It reflects an appropriate response to evolving financial markets,[72] in particular, the fact that the interests of creditors and shareholders have changed since the era during which the theories supporting the mandatory subordination of shareholder claims were developed.[73] It is worth noting that more than a quarter of a century ago, the argument was already being advanced that the sophistication in financial law and practice that had evolved since the inception of the subordination doctrine severely undermined the policy basis for protecting the creditors.[74] Shareholders were seen as constituting a more broadly dispersed group and no longer just a minor assembly of entrepreneurs and local investors, while creditor groupings had yielded to the dominance of financial institutions in place of comparatively small trade creditors and individual debt holders.[75] Both modern business creditors and modern shareholders are seen to have greater resources at their disposal to evaluate a corporation's financial position, and a stronger understanding of the factors that affect solvency than they had when mandatory subordination was first introduced.[76] Thus, the comparative abilities of the debt and equity classes to protect themselves from fraud and to represent their interests vigorously in a bankruptcy proceeding may be taken to have changed.[77] On this view, there is no conflict between the absolute priority rule and the operation of the Fair Funds provision in a bankruptcy. Not only is the result entirely in accord with the reality of the circumstances of creditors and shareholders in current times, it is argued that it also

70 See Explanatory Note to the Corporations Amendment (No 2) Bill 2010, paras 2.6–12.

71 Henry, "Subordinating subordination", n. 24 above, at p. 262.

72 Ibid.

73 Ibid. p. 284. See Duffy's discussion of the core issues in *Sons of Gwalia* from the perspective of the "stakeholder" and "nexus of contracts" theories of the corporation: M Duffy, "After *Sons of Gwalia* – some perspectives on the position of shareholders and creditors and the question of law reform", (2008) 22 *Australian Journal of Corporate Law* 161.

74 Davis, "The status", n. 15 above, pp. 28–9.

75 Ibid; Henry, "Subordinating subordination", n. 24 above, at p. 284.

76 Henry, "Subordinating subordination", n. 24 above, at p. 284.

77 Ibid; Davis, "The status", n. 15 above, p. 29, in fact argued that the *shareholders* had limited means to acquaint themselves with the debtor's financial situation. Likewise, the CAMAC Report, n. 69 above, notes that in practice, existing shareholders may have little or no real day-to-day capacity to monitor or control corporate disclosures or other corporate conduct and may be as misled as new investors by corporate misconduct (para. 3.4).

represents an acceptance that blanket shareholder subordination is a rigid approach which can yield unfair outcomes.[78]

Moving beyond the legal position in the US, there are further broad justifications which may be identified in favour of elevating shareholder claims. The first of these is that the inherent dynamism of public policy must extend to a rule such as *pari passu*. That is to say, since *pari passu* has been upheld and applied as being based on public policy,[79] and public policy is based on the current needs of the community,[80] in principle the concept of rateable treatment should be equally malleable to accommodate the effect of the changes in the relationship between shareholders and creditors in modern times. Indeed, it was acknowledged, among the arguments for non-subordination preceding the recent Australian Corporations Amendment Bill, that the positions of shareholders and creditors in an insolvency are not wholly dissimilar:

> [I]n some companies, such as large listed companies, ordinary shareholders, even institutional shareholders, have limited practical ability to direct the company and in reality may have no greater power than creditors. They therefore need a comparable level of protection in an insolvency.[81]

Since the law develops by perpetually drawing new values and solutions from the life of the community[82] – a feat attained partly through the development of new law, and partly through standards and principles which are implicit in particular branches of the law such as "reasonableness" and "public policy"[83] – a change of this type is not beyond the realm of possibility. Moreover, it has been argued that some of a company's securities violations may have occurred with the collusion of some of its creditors (albeit that their actions may not attract specific civil or criminal liability);[84] and since trade creditors dealing with such a company will typically have their invoices paid until the company is on the verge of insolvency, they are not invariably in the position of parties who have lost everything to a debtor.[85] In the light of all these factors, it would seem inconsistent with the development of public policy that mandatory subordination should override the possibility of any payment to defrauded shareholders on the same priority with the unsecured creditors of a company. It is arguably desirable that the understanding of "equality of distribution" should evolve in tandem with this re-balancing of the relative positions of creditors and shareholders. Indeed, remaining faithful to the theoretical conception of the risks assumed by either grouping would achieve the result that the law pulls in a different direction from the realities of commerce and, thus, fails to meet the broad goal of public policy to serve the current needs of the community.

Another factor which should enable the acceptance of *pari passu* treatment for defrauded shareholders vis-à-vis unsecured creditors is the notion that companies are social

78 A Hargovan and J Harris, "*Sons of Gwalia* and statutory debt subordination: an appraisal of the North American experience" (2007) 20 *Australian Journal of Corporate Law* 265.

79 See *British Eagle* case, n. 56 above; and *Attorney General* v *McMillan & Lockwood* [1991] 1 NZLR 53. See, however, more recent comments narrowing the scope of this public policy aspect of the rule in *Perpetual Trustee* and *IATA* v *Ansett*, both cited in full at n. 56 above.

80 P H Winfield, "Public policy in the English common law" (1928) 42 *Harvard Law Review* 76–102, p. 92.

81 CAMAC Report, n. 69 above, para 3.3.1. See also Explanatory Note, n. 70 above, para. 2.23.

82 G Paton, *Jurisprudence* (Oxford: Clarendon 1972), p. 199.

83 Ibid.

84 K Cordry, "Categorical subordination: still kicking?" (2006) 24(10) *American Bankruptcy Institute Journal* 8, p. 51.

85 Ibid.

enterprises[86] – their existence is required to produce consequences beneficial to society[87] and their activities should accordingly be consonant with the public interest.[88] This includes making profits for the shareholders, which becomes a mechanism for promoting the public interest, and not an end in itself.[89] The classification of companies as social enterprises entitles the state to intervene in order to safeguard the public interest and ensure compliance with publicly acceptable ethical standards while they are going concerns.[90] On this basis, it is submitted that this entitlement should reasonably continue into the insolvency of a company and would be even more pertinent to a reorganisation process which offers the prospect of survival of a company. The public interest receives protection through the deterrence of fraud or dishonest behaviour on the part of companies, the preservation of financial integrity and commercial morality in the market, and the encouragement of investment or enterprise within society. These concerns were prominent in the debates leading to the enactment of the Sarbanes–Oxley Act[91] and, from an enforcement perspective in particular, actions such as disgorgement and restitution to injured investors promote economic and social policies, including investor confidence in the fairness and transparency of securities markets and the deterrence of future violations.[92] It is similarly noted by the Australian legislature that "aggrieved shareholder claims" can act as a form of private enforcement and help promote the integrity of corporate conduct, to the benefit of lenders and the market generally, and not only shareholders.[93] This resonates with the view that from a social-enterprise perspective financial impropriety on the part of managers and inadequate methods of accountability and control are no more to be tolerated in the corporate sphere than they are within the organs of government.[94] Thus, to accept that insolvency law affects community interests – and is bound to recognise and safeguard such interests[95] – is to extend the social-enterprise/public-interest aspect of a company's existence into its demise. A virtuous circle is identifiable between the efficiency

86 J E Parkinson, *Corporate Power and Responsibility: Issues in the theory of company law* (Oxford: Clarendon 1993), p. 23. See also Armour et al.'s submission that the goal of corporate law is to advance overall social welfare: J Armour, H Hansmann and R Kraakman, "What is corporate law?" in R Kraakman, J Armour, P Davies et al., *The Anatomy of Corporate Law: A comparative and functional approach* 2nd edn (Oxford: OUP 2009), ch. 1, p. 28.

87 Parkinson, *Corporate Power*, n. 86 above, p. 33.

88 Parkinson, *Corporate Power*, n. 86 above, pp.23–6.

89 Ibid. p. 24.

90 Ibid. In similar vein, Cheffins finds support for the proposition that lawmakers should take action when markets threaten to undermine key community ideals: B Cheffins, *Company Law: Theory, structure and operation* (Oxford: Clarendon 1997), p. 153.

91 148 Cong Rec H4478-85 ("Corporate reform needed" – House of Representatives); 148 Cong Rec H5462-80 (conference report on proposed legislation, House of Representatives); and 148 Cong Rec S7350-65 (conference report on proposed legislation, Senate).

92 SEC, *Report Pursuant to Section 308(c) of the Sarbanes–Oxley Act of 2002*, available from www.sec.gov/news/studies/sox308creport.pdf , pp. 19–20.

93 See Explanatory Note, n. 70 above. This echoes the CAMAC Report's (n. 69 above) position: "Claims by aggrieved shareholders can serve as a market-based deterrence, enforcement and recovery mechanism in support of required standards of corporate conduct" (para. 3.4).

94 Parkinson, *Corporate Power*, n. 86 above, p. 202. Indeed, Attenborough notes that the very largest corporations have greater economic power than many nation states: D Attenborough, "How directors should act when owing duties to the companies' shareholders: why we need to stop applying Greenhalgh" (2009) *International Company & Commercial Law Review* 339, p. 346.

95 I*nsolvency Law and Practice: Report of the Review Committee*, Cmnd 8558 (1982), paras 198(i), 240 and 1734.

of financial markets resulting from timely and accurate disclosures,[96] and a reduced risk of insolvency as fewer companies fail through poor management and the delayed evidence of financial difficulty.

Insolvency law jurisprudence[97] and policy[98] do not deny the relevance of public interests, but do not answer fully the important question of the extent to which they can (or should) be accommodated in practice. This has provided fertile ground for debate, in particular as regards the ability of social interests to impinge on creditors' rights of recovery.[99] However, in the operation of Sarbanes–Oxley Fair Funds provisions, there is no direct encroachment on the assets available to unsecured creditors since the SEC recovers restitution by way of civil penalties and disgorgement levied on a wide body of persons and pays to the injured investors out of funds that would otherwise be paid to the Treasury.[100] The new position is therefore arguably defensible on the ground that it does not interfere with the equal treatment of creditors per se, as the disbursements are not being made from funds that they would otherwise be entitled to. In fact, in some cases, the SEC has made payment of disgorgement or civil penalties to investors via a bankruptcy trustee "for distribution to creditors, including investors".[101] Although this particular parity may seem inconsistent with the distribution structure enshrined in the absolute priority rule, in reality the scope for interference with this structure is reduced by factors pertinent to the operation of Fair Funds. These include the difficulty faced by the SEC in satisfying strict judicial requirements for disgorgement orders;[102] the power of the SEC to exercise a discretion to either transmit its recoveries to the Treasury or use them to compensate the defrauded investors;[103] the likelihood that such distributions may be foregone in situations where it proves not to be cost-effective to make them;[104] and the principle that compensation is not the same as satisfaction of the shareholders' claims.[105] The impact of the application of Fair Funds on the maintenance of the insolvency distribution structure may thus not be as profound as it might appear at first glance. It is accordingly proposed that for insolvency law to be seen to do more than pay lip-service to notions of the community welfare, it should give perceptible effect to such market-related concerns as those targeted by the Sarbanes–Oxley Act, particularly if this offers the prospect of reducing the incidence of insolvency in those same markets. Moreover, it would show that the demands of commercial morality placed upon a company during its lifetime (as

96 CAMAC Report, n. 69 above, para. 2.4.3: "Financial markets are more efficient and less volatile to the extent that companies provide timely and accurate disclosures about their real financial position and prospects."

97 See, for instance, A Keay, "Insolvency law: a matter of public interest?" (2000) 51 *NILQ* 509–34; and V Finch, *Corporate Insolvency Law: Perspectives and principles* 2nd edn (Cambridge: CUP 2009), ch. 2.

98 *Insolvency Law and Practice*, n. 95 above, paras 198(i), 240 and 1734; Law Reform Commission, *General Insolvency Inquiry*, Report No 45 (Commonwealth of Australia, 1988), para. 33.

99 The various views have been outlined in some detail by Finch, *Corporate Insolvency Law*, n. 97 above.

100 See SEC, *Report*, n. 92 above, pp. 3–5. See also Christensen, "Fair Funds", n. 26 above, p. 371.

101 SEC, *Report*, n. 92 above, pp. 10, 12 and 16.

102 Ibid. pp. 18–19 and 23–4. A reading of the Fair Funds provision shows that it cannot operate in the absence of a disgorgement order, to which the SEC may add civil penalties. There is a danger of a practice developing whereby the SEC seeks token disgorgement amounts in order to facilitate the distribution of massive corporate penalties to investors: B Black, "Should the SEC be a collection agency for defrauded investors?" (2008) 63 *Business Lawyer* 317, at p. 330.

103 SEC, *Report*, n. 92 above, pp. 5 and 10–11.

104 Ibid. pp. 3–5.

105 *SEC* v *Fischbach Corporation* 1333 F3d 170; and SEC, *Report*, n. 92 above, pp. 19–20. Winship, "Fair Funds", n. 33 above, points out that the losses sustained by investors often dwarf the profits made by the violators and thus recoveries only represent a fraction of the amounts lost (5–6% in the case of WorldCom), p. 1125.

evidenced by measures including the Sarbanes–Oxley Act)[106] are accorded the same significance on its insolvency.

Focusing more narrowly on debtor/creditor and inter-creditor relationships, one might also contend for an acceptance of the improved position of shareholders relative to unsecured creditors on the ground that creditors generally have a number of alternatives open to them for the purpose of protecting themselves from the consequences of a debtor's insolvency. They are often able to engage in processes to determine the probability of the debtor's default in advance; they may obtain personal guarantees or security; require information which demonstrates the debtor's continuing creditworthiness, and take insurance against the risk of default.[107] It is noted that apart from their ability to negotiate for contractual protections, many creditors are well-diversified, in the sense that each debt contract only has a small impact on their financial status.[108] A typical lender is ordinarily well-situated to absorb the loss associated with the failure of a single business enterprise, whether this lender takes the form of a trade creditor with numerous customers or a bank with a large number of corporate borrowers.[109] The historical necessity for granting rigid creditor protections as a quid pro quo for the limited liability enjoyed by shareholders is therefore distinguished from modern commercial practice, where major business creditors rely not on the law, but on contract, credit agencies and a host of other self-help measures to safeguard their interests.[110] The reach of creditor influence is evidenced by the fact that much of the concern regarding the implications of the *Sons of Gwalia* decision was directed towards the effects on debt finance for companies. It was perceived that there would be a reduced availability or increased cost of finance and lenders would be more likely to seek security or guarantees and impose restrictive conditions on loans.[111] Trade creditors would have increased recourse to retention of title agreements, or factor the added risk of non-recovery in insolvency into their costs of goods or services.[112] It followed, therefore, that explicit statutory subordination of shareholder claims would facilitate the provision of credit to companies and, from the standpoint of credit providers, reduce risk premiums and the imposition of onerous terms and conditions.[113]

In addition, as between creditors *inter se*, it is evident that many of the rules of insolvency distribution governing their relationship are already underpinned by notions of fairness. Drawing from the example of English insolvency law, in the validation of pre-insolvency dispositions, transactions which might result in certain creditors being paid in full at the expense of other creditors who will only receive a dividend may be upheld where special circumstances exist making such a course desirable in the interests of the unsecured creditors as a body.[114] Similarly, a preferential transaction brought about by proper

106 See further Cheffins, *Company Law*, n. 90 above, discussion on the ways in which the state will intervene to ensure fairness in the treatment of company participants, effective participation in corporate affairs, the protection of community ideals, and the preservation of morality in the market system, pp. 142–58.

107 Cheffins, *Company Law*, n. 90 above, pp. 501–2. A view also shared by Davis, "The status", n. 15 above.

108 Cheffins, *Company Law*, n. 90 above, p. 502.

109 Ibid.

110 Armour et al., "What is corporate law?", n. 86 above, make this observation in advancing the argument that corporate law is useful as a foundation and supplement to contractual protections for creditors, at p.72.

111 CAMAC Report, n. 69 above, para. 2.4. See also paras 2.6–9 of Explanatory Note, n. 70 above.

112 CAMAC Report, n. 69 above, para. 2.4.

113 Explanatory Note, n. 70 above, para. 2.12.

114 *Re Gray's Inn Construction Co. Ltd* [1980] 1 WLR 711, at 717D and 718 A–C; *Re J Leslie Engineers Co. Ltd (in liq)* [1976] 1 WLR 292, at 305; *Denney* v *John Hudson & Co. Ltd* [1992] BCLC 901, at 904; and *Re Tain Construction Ltd* [2003] EWHC 1737, paras 14–17.

commercial considerations[115] or a late floating charge granted to creditors involved in efforts to revive a struggling company[116] have been countenanced, notwithstanding the consequences for equal treatment. Defences based on the "ordinary course of business", entitling creditors to retain benefits from preferential transactions, are also a feature of US and Australian insolvency law.[117] If these differences among creditors may be accepted on the grounds of fairness, strong support may be lent to a claim by injured shareholders to similar treatment on the grounds of procedural fairness.[118] Procedural fairness relates to the fairness of the contracting process,[119] encompasses the methods which market participants use to negotiate and formulate transactions, and is breached where the transactors do not have a chance to make agreements freely and knowingly.[120] Where shareholders have purchased shares on the basis of fraudulent misrepresentations, there is a clear absence of procedural fairness. Injured investors are required to prove the fact of the dishonest inducement, together with their dependence thereon and consequent loss, to the satisfaction of a court in order to be awarded damages[121] – a burden conceded to be a difficult one to discharge, in the analysis following *Sons of Gwalia*.[122] It is therefore arguable that their entitlement to fairness (connoting similar treatment to unsecured creditors) is judicially established in comparison to the notion of creditor reliance on capital reserves, which exists as a mere presumption of doubtful influence.[123] In addition, to the extent that the concept of fairness is taken to encompass a moral element,[124] it would seem even more important that it should avail relief in the case of involuntary debt (that is, the defrauded shareholders) compared with fairness based on commercial expediency (to wit, the general creditors in whose favour normally voidable transactions are adjusted). Extending this equitable treatment to accommodate the defrauded shareholders thus accords with the same principle which enables creditors to adjust their own relationships, and would not significantly distort existing insolvency processes or relative creditor positions within the unsecured rank. Furthermore, it can be reconciled with an important rule of equity, namely that, where trustees have made a fraudulent conveyance, the loss should fall on the beneficiaries rather than on a bona fide purchaser for value.[125]

115 *Re MC Bacon Ltd* [1990] BCLC 324, at 336; *Lewis* v *DKG Contractors* [1990] BCLC 903, at 910; *Re Fairway Magazines Ltd* [1992] BCC 924, at 930; *Wills* v *Corfe Joinery Ltd* [1998] 2 BCLC 511, at 512.

116 Insolvency Act 1986, s. 245(2)(b); *Re Matthew Ellis Ltd* [1933] Ch 458.

117 US Bankruptcy Code, §547(c); Corporations Act 2001 (Australia), s. 588FG(2).

118 Described by S Smith, "In defence of substantive fairness" (1996) 112 *Law Quarterly Review* 138, p. 140; and Cheffins, *Company Law*, n. 90 above, p. 143.

119 Smith, "In defence", n. 118 above.

120 Cheffins, *Company Law*, n. 90 above; Smith, "In defence", n. 118 above.

121 Under contract and traditional tort law principles: Davis, "The status", n. 15 above, p. 39. Reliance is presumed, under the "fraud on the market" theory in the US – see *Basic* v *Levinson* (1988) 485 US 224. See also para. 3, Sch. 10A of the UK Financial Services and Markets Act 2000.

122 CAMAC Report, n. 69 above, at para. 3.2.2: obtaining a remedy through litigation as an aggrieved shareholder can be a difficult task, as it turns on whether a shareholder can establish the necessary elements of relevant corporate misconduct, causation, reliance and damages incurred.

123 Davis, "The status", n. 15 above, at pp. 32–4, questions the extent to which the existence of the subordination doctrine influences lending decisions. The High Court of Australia in *Sons of Gwalia* [2007] HCA 1 also expressed doubts as to whether the reliance principle "reflects the reality of modern commercial conditions, where assets and liabilities usually are more significant for creditors than paid-up capital" (para. 5).

124 D Sullivan, "Rules, fairness, and formal justice" (1975) 85 *Ethics* 322–31. Admittedly, however, this is unfair on the remaining shareholders: the CAMAC Report notes that when aggrieved shareholders sue or reach a settlement with the company, the people who indirectly suffer loss are the remaining shareholders, from the market value of their shares or reduced dividends (n. 69 above, at para. 3.3.4). This problem is discussed further below.

125 *Pilcher* v *Rawlins* (1872) 7 Ch App 259.

Difficulties with implementing parity

It may be seen from the foregoing that, on the basis of corporate theory and the nature of public policy, there are grounds for considering that the US legislature is content with the alignment of the interests of unsecured creditors and injured investors and might not respond to calls for it to resolve the apparent discord between the principle of mandatory subordination and the effect of a Fair Funds distribution.[126] This may be contrasted with the swift action taken in Australia to enact an unequivocal rule. Many of the concerns voiced relate to the manner in which the SEC or injured investors may adapt their behaviour to maximise their recoveries.

For example, fears have been expressed that the SEC may become aggressive in its pursuit of larger penalties, resulting in the decrease of assets available to general creditors, where the penalties are sought from a company involved in a bankruptcy proceeding.[127] These are reinforced by indications that the SEC has welcomed the Fair Funds provision as an innovation which it can use "to return more funds to investors".[128] Black relies on financial fraud settlements against four corporate defendants to demonstrate how the SEC's efforts to create Fair Funds distributions for investors have resulted in an evasion of the disgorgement requirement and the imposition of sizeable penalties.[129] Moreover, the shift towards compensation as a goal of the SEC may have implications for the deterrence objective of securities law enforcement, if the SEC pursues penalties that have a compensatory effect but are inadequate for deterrence, or foregoes the collection of penalties in situations where they cannot be used to compensate investors.[130]

Mass claims by investors to recover damages may also bring about a diminution in the assets available to unsecured creditors, as noted by Callinan J in *Sons of Gwalia*: "It is not difficult to imagine a situation in which claims of a large body of shareholders, perhaps most of them, would dilute the creditors' rights to a trickle."[131]

In the context of Australian law, it was noted that class actions and litigation funding might encourage this form of shareholder litigation, also to the detriment of the remaining shareholders.[132] Furthermore, any complications associated with such proceedings would be multiplied in the event of the collapse of a corporate group.[133]

A matter which has received less attention, particularly in the context of the US where the SEC as a regulatory body is empowered to redistribute its bankruptcy recoveries to injured investors, is the extent to which these "sub-distributions" conform to the principles of insolvency distribution. The potential for the SEC, in the context of bankruptcy proceedings, to pursue policies which are not related to the accepted objectives of insolvency law[134] is most starkly reflected against the backdrop of the three possible

126 Made by scholars including Christensen, "Fair Funds", n. 26 above.

127 Ibid. p. 370.

128 SEC, *Report*, n. 92 above, p. 36. See Winship, "Fair Funds", n. 33 above, p. 1124.

129 Black, "Should the SEC?", n. 102 above, p. 331. Namely WorldCom, Time Warner, AIG and Fannie Mae.

130 Winship, "Fair Funds", n. 33 above, at p. 1139.

131 *Sons of Gwalia* [2007] HCA 1, para. 256.

132 See CAMAC Report, n. 69 above, n. 3.3.4.

133 M Wyburn, "Pooling as a response to the competing interests in corporate group collapses in Australia" (2010) 19 *International Insolvency Review* 65, p. 72.

134 Objectives outlined by Goode, *Principles*, n. 12 above, pp. 2–17 *et seq*; Richardson and Mack note that the SEC can establish a plan based on considerations that are not recognised under the US Bankruptcy Code: R G Richardson and J S Mack, "It isn't what it used to be, but the SEC still protects shareholder interests" (2007) 26(3) *American Bankruptcy Institute Journal* 12, p. 67.

interpretations that are associated with the mention of the *pari passu* principle,[135] which enforces pro rata distribution among unsecured creditors in insolvency. This "equality of distribution" principle has been identified in the opening pages of this paper as operating within the framework of absolute priority in US law, and accordingly may be examined within the setting of more general observations relating to *pari passu* treatment.

The broadest view of the *pari passu* rule is one whereby it connotes the treatment of all unsecured creditors equally and ignores any considerations pertaining to their individual circumstances.[136] This outlook is not represented in the SEC's actions vis-à-vis Fair Funds because it is empowered to set distribution criteria that are not objective in nature. The courts, recognising the deterrent role of enforcement mechanisms such as disgorgement, have deferred to "the experience and expertise of the SEC" in the line-drawing which inevitably leaves out certain potential claimants.[137] This line-drawing is as relevant for seeing which parties are brought into the distribution process, as well as noting who is left out: as seen in *Official Committee of Unsecured Creditors of WorldCom Inc.* v *SEC*,[138] shareholders may receive payment in priority to certain classes of creditors in the interests of equalising their respective recoveries. This means that pre-insolvency payments to creditors, which would not otherwise be subject to challenge, may nonetheless provide grounds for excluding them from the SEC's distribution. Furthermore, since the Fair Funds proceeds may be distributed among "victims of [securities] violations"[139] it is possible for parties who had not acquired an interest in the debtor's estate as shareholders, but were nonetheless injured by the violations, to enjoy priority over non-defrauded shareholders. This goes to the nature of the rights acquired before bankruptcy[140] and, with respect to personal actions, the upholding of remedies in tort over the contractual rights of creditors and shareholders.

An alternative conception of *pari passu* treatment sees it as denoting no more than pro rata distribution within classes.[141] This is detracted from in the context of Fair Funds insofar as experience has shown that the SEC may choose which parties to exclude in favour of others. In terms of this second interpretation, in the normal bankruptcy hierarchy, non-preferential unsecured creditors would together occupy one class and shareholders another.[142] By contrast, the SEC's distribution methods permit it to select from either class certain investors to benefit from the Fair Funds proceeds. This would be at the expense of others who properly belong to those classes but who had made a profit on the sale of their securities during the period in which the fraud occurred or recovered more than a given percentage of their entitlements through the sale of their securities.[143] Even though these investors could have qualified as general unsecured creditors or shareholders, they may be barred by the SEC from participation in the Fair Funds distribution for the benefit of the non-profiting or non-recovering unsecured creditors and shareholders. The result is

135 Identified by R J Mokal and L C Ho, "The *pari passu* principle in English ancillary proceedings: *Re Home Insurance Company*" (2005) 6 *Insolvency Law and Practice* 207–10, p. 208.

136 Mokal and Ho, "The *pari passu* principle", n. 135 above; Bridge, "Collectivity", n. 21 above, para. 1.2; *In re Smith, Knight & Co. ex parte Ashbury* (1868) LR 5 Eq 223.

137 *SEC* v *Wang* (1991) 944 F2d 80, at 88.

138 *Official Committee*, n. 42 above.

139 §308(a) Sarbanes–Oxley Act, 2002 Pub L No 107-204, 116 Stat 745.

140 Goode, *Principles*, n. 12 above, para. 3-02.

141 Mokal and Ho, "*Pari passu* principle", n. 135 above.

142 Ibid; Wood, *Law and Practice*, n. 11 above, para. 1-14.

143 Distributions of this nature accepted as being proper in *SEC* v *Wang* (1991) 944 F2d 80 and in *Official Committee*, n. 42 above.

therefore of differing treatment of general creditors *inter se* and shareholders among themselves by effecting the dismantling of these recognised classes in insolvency. If we accept that ranking in insolvency is indeed its most important feature, rather than the achievement of equal treatment,[144] in the context of the sub-distributions creditors or shareholders cannot be guided by the prior knowledge that they would ordinarily have of their place within the distribution structure. Thus, their rank (or mere participation in distribution) is not determined with reference to negotiations conducted with the debtor *ex ante*[145] and in accordance with their respective bargains, but in the light of their position at the time of distribution and *relative to one another*. Insofar as account is taken of profits or recoveries already made by the time the SEC comes to disburse Fair Funds proceeds, the ranking is coloured with a moral or ethical element which *pari passu* has been found to lack.[146] This disregards the possibility that at least some of the transactions may have been entered into by the investors in ignorance of the fraud. Even where there is an awareness of the fraud, it might not necessarily act as a check on pre-disgorgement transactions between the debtor and investors: some may prefer to maximise the gains from the sale of their securities than gamble on the uncertainty of the SEC's discretion being exercised in their favour. Thus, rather than producing deterrent effects for irregular or dishonest activities, this form of ranking may be counter-productive for the SEC's pooling of funds.

The entitlement of the SEC to distribute Fair Funds proceeds according to its discretion also sits uneasily with the third facet of *pari passu* treatment, namely the principle of collectivity.[147] The collective nature of insolvency relates to the conservation of the estate to ensure orderly distribution among creditors.[148] Predictably therefore, the number of ordinary unsecured creditors proving claims in the debtor's insolvency determines the size of their respective returns following payment of priority obligations. Collectivity consequently carries a certainty of at least partial recovery for a creditor, with the proportion being dependent upon the number of other parties participating in the sharing of the estate. Conversely, in the disbursement of Fair Funds, it is settled that the standard of fairness and reasonableness permits the SEC to take account of the limited funds available for distribution in deciding which parties to exclude from payment under its plan.[149] Thus, the certainty of participation is absent for investors as the amount available for payment influences the decision on who may benefit from the Fair Funds. The distinction is clear: while for the purposes of collective treatment in insolvency it is sufficient to be accepted as a creditor, qualifying as a victim of a securities violation does not assure enjoyment of a share in Fair Funds.

A gap is therefore apparent between the normal course of an insolvency distribution procedure and the manner in which the SEC fulfils its responsibilities for distribution of the Fair Funds with respect to a company which is now insolvent. The readiness of other jurisdictions to adopt the Fair Funds model may be tempered by the fresh imbalances which parity seems to generate. These are briefly discussed in the concluding section of this paper.

144 J Dalhuisen, *Dalhuisen on Transnational and Comparative Commercial, Financial and Trade Law* 3rd edn (Oxford: Hart 2007), p. 959. See also Mokal and Ho, "*Pari passu* principle", n. 135 above.

145 As argued by Dalhuisen, *Dalhuisen*, n. 144 above.

146 See, for instance, *Mitchell and another* v *Buckingham International plc & others* [1983] Ch 1. "Amorality" of *pari passu* also noted by Bridge, "Collectivity", n. 21 above, para. 1.2.

147 Principle outlined in more detail by Mokal and Ho, "*Pari passu* principle", n. 135 above, p. 208.

148 Ibid. Goode, *Principles*, n. 12 above, para. 1-04.

149 *SEC* v *Wang* (1991) 944 F2d 80, at 87–88. This also occurred in *Official Committee*, n. 42 above. This is despite the fact that the SEC can determine the size of its "estate" insofar as it can take some account of shareholder loss in formulating the size and nature of its penalty: *SEC* v *WorldCom* 273 F Supp 2d 431.

Does parity open up new imbalances?

The difficulty of squaring the parity brought about by Fair Funds with prevailing conceptions of the notion of *pari passu* treatment have been highlighted in the preceding section of this paper. Even if it could be aligned with the equality of distribution principle, there are two important respects in which it may be seen as unsettling particular shareholders and creditors.

The first, identified in literature on securities class actions, is the circularity problem.[150] This occurs due to the shifting of wealth from the current shareholders of a corporation, who indirectly bear the costs of any judgment or settlement against it, to the claimant shareholders who acquired shares in the company at the material time.[151] Coffee describes a more complex form of wealth transfer, affecting diversified shareholders, in these terms:

> Often shareholders will belong to both the plaintiff class that sues and the residual shareholder class that bears the cost of the litigation. This can result because they purchased stock at times that are both inside and outside the class period, so that they are on both sides of the litigation. Thus, they are effectively making wealth transfers to themselves, in effect shifting money from one pocket to the other, minus the high transaction costs of securities litigation.[152]

Alternatively, if the shareholders of a company are not all similarly diversified, litigation may bring into tension the interests of the "buy and hold" (small undiversified) investors with those of the "in and out" (larger sophisticated) traders.[153] Mitchell has decried the notion of the "innocent shareholder", portrayed as the rationally apathetic passive investors from whom wealth is transferred in this way. He argues that shareholders' participation in the affairs of the company ensures the integrity of capital markets and is part of the mechanism by which managerial frauds are deterred.[154] This may be countered with the observation that shareholders of a public corporation usually have little or no voice in the selection of the managers or the way in which they conduct its affairs; large creditors are likely to have much more control, especially as the company slides into insolvency.[155] One might also add that the "shareholder participation" argument does not answer the concerns of individuals who hold shares indirectly through pension funds, unit trusts, or other collective investments.[156] They are not in a position to exert any significant influence over the running of the company.

150 J Coffee, "Reforming the securities class action: an essay on deterrence and its implementation" (2006) 106 *Columbia Law Review* 1534; L Mitchell. "The 'innocent shareholder': an essay on compensation and deterrence in securities-class actions" (2009) *Wisconsin Law Review* 243.

151 Coffee, "Reforming", n. 150 above.

152 Ibid. p. 1558.

153 Ibid. pp. 1559–60.

154 Mitchell, "Innocent shareholder", n. 150 above, pp. 291–2.

155 Davis, "The status", n. 15 above, p. 44. Furthermore, with respect to individual shareholders, the evolution of their role from "shareholder as owner/principal" to "shareholder as investor" with limited participation rights is traced by Jennifer Hill: J Hill "Visions and revisions of the shareholder" (2000) 48 *American Journal of Comparative Law* 39.

156 A recent report on equity and bond ownership in the US found that that there had been a growth in ownership through employer-sponsored retirement plans: *Equity and Bond Ownership in America* (Washington: Investment Company Institute and the Securities Industry and Financial Markets Association 2008).

Secondly, the debate regarding parity creates a danger of viewing all unsecured creditors as contractual creditors, such as lenders and trade suppliers.[157] Parity will also affect creditors who have not bargained for risks on the same basis. While parity with non-shareholder tort creditors may be wholly appropriate on the basis that both types of claim are rooted in involuntary debt, in this context one must also have regard to the interests of creditors, such as the revenue authorities, whose relationship with the debtor company is not derived from a negotiated acceptance of risks. This is of particular significance in countries where fiscal debts no longer enjoy preferential status, such as the UK. Insolvency law reforms aimed at improving the position of unsecured creditors included the abolition of Crown preference in all insolvencies – corporate and personal.[158] A study recently carried out by the Office of Fair Trading found that Her Majesty's Revenue & Customs was owed more than £50,000 in 57 per cent of administrations and on average accounted for 24 per cent of unsecured debt.[159] The justifications for parity may be somewhat undermined, if it results in encroachment on other claims which merit repayment in the wider public interest.

The potential for these distortions to result from an acceptance of parity underlines the importance of clarifying who the "shareholders" and "unsecured creditors" are. That is to say, it is difficult to assess the weight to be given to arguments for parity without knowing whether the actual composition of these two groupings substantiates perceptions of their relative power or vulnerability. No statistical breakdowns could be found regarding the make-up of the unsecured creditor class,[160] but studies of share ownership in the US and Australia respectively reveal a great deal about the nature and extent of share investments in the two countries. It is recorded that in the US in 2008, nearly half of all households owned either equity or bonds.[161] Depending on their age, the primary goals of investors were saving for a home purchase and education (under 40s), saving for retirement (40–64 years) or generating current income (65 years and older).[162] Having long-term investment goals meant that most equity-owning households were not frequent traders, and had not evidenced a pattern of buying and selling in response to the stock market conditions in 2007.[163] These appear to be the buy-and-hold investors described in the wealth transfer problem mentioned above, who indirectly bear the cost of judgments or settlements against the company in favour of shorter-term investors. Furthermore, it was found that a large number of investors had purchased their equities/bonds through professional financial

157 For instance, the CAMAC Report, n. 69 above, in Australia considered the implications for corporate financing (domestic and international), trade creditors, and the standing of Australian companies in corporate bond and distressed debt markets, paras 2.4.1–3. Davis, a proponent of parity in the US long before the introduction of Fair Funds, also sees the creditor class as being dominated by large financial institutions "although it includes many comparatively small claimants in the form of trade creditors and individual holders of the debtor's bonds and debentures"; Davis, "The status", n. 15 above, p. 29.

158 See para. 2.19 of the government White Paper, *Productivity and Enterprise: Insolvency: A second chance* CM 5234 (London: HMSO/DTI 2001). Abolished via s. 251 of Enterprise Act 2002, amending Sch. 6 of the Insolvency Act 1986 to remove preference for debts due to the Inland Revenue, Customs and Excise, and in respect of social security contributions.

159 Office of Fair Trading, *The Market for Corporate Insolvency Practitioners* (London: Office of Fair Trading/Crown 2010), para. 4.61.

160 The study by the UK Office of Fair Trading states that "[u]nsecured creditors range from larger, repeat unsecured creditors such as Her Majesty's Revenue & Customs . . . and large landlords (in England & Wales) to customers, trade creditors and employees"; ibid. para. 1.13.

161 *Equity and Bond Ownership*, n. 156 above, p. 5.

162 Ibid. pp. 27–8.

163 Ibid. p. 47.

advisers, rather than through retirement plans at work.[164] Contrary to the perception of modern shareholders being resourceful and more knowledgeable about the risks attached to different share offerings,[165] ownership of equities through professional advisers was predominant across all investor groups, regardless of age, education level, household income, etc.[166] Most households that acquired equities through professional advisers were shown to regularly rely on their advisers for investment advice and guidance.[167] Likewise, in Australia 41 per cent of the adult population owned shares in 2008, the main reasons for investing being "to make money"; accumulate wealth; for long-term capital gains; and to obtain higher returns.[168] There was no marked pattern of frequent buying and selling: between 2002 and 2008 the average value of share trades did not exceed 11 per cent of the average value of monies invested.[169] Australian investors could be distinguished from those in the US by their use of the internet to buy shares, and the division of share-owners into segments according to their level of knowledge/skill and passion for investing.[170] A sharp contrast may be drawn between the two countries and the UK where individual share ownership levels have declined.[171]

The recognition that a significant number of investors are private individuals seeking to raise funds to pay for their homes, make provision for education or secure their retirement[172] therefore adds another dimension to this debate. These parties may be deemed to be in a comparable position to that of unsecured creditors, in that they are in truth dealing with the company as outsiders, with similar limitations in their knowledge of its business and matching expectations to be supplied with accurate information. The studies detailing their long-term outlook on investment, infrequent trading activity, and limited self-reliance in transacting on the market, indicate that their insight into the risks attached to their investments is not as keen as it may be deemed to be.

Conclusion

This paper has sought to demonstrate how, as a matter of principle, grounds exist for accepting parity as a response to modern conditions. Nevertheless, it is clear that the manner and consequences of implementing such a policy require careful analysis. In

164 *Equity and Bond Ownership*, n. 156 above, p. 37.

165 See text accompanying nn. 75–7 above.

166 *Equity and Bond Ownership*, n. 156 above, p. 45.

167 69% of these equity/bond owners "always" or "sometimes" consult their advisers when making investment decisions: ibid. p. 41.

168 ASX, *2008 Australian Share Ownership Study* (Sydney: Australian Securities Exchange 2009), pp. 3–15.

169 Ibid. p. 18.

170 56% of investors bought shares through an online broker in 2008 (ibid p. 20), compared with the US where the least frequently-mentioned use of the internet was to buy or sell stocks/bonds (see *Equity and Bond Ownership*, n. 156 above, p. 37) – investors used it mainly to access financial accounts or get financial news. As regards segmentation, ASX 2008 classified investors as "confident", "aspirational", "diligent" or "delegator".

171 10.2% of the value of all UK ordinary shares quoted on the London Stock Exchange is held by individuals, compared with 54% in 1963 and 14.1% in 2001: see Office for National Statistics, *Share Ownership Survey* (London: Crown 2010). In terms of household percentage, 34.2% of UK households hold shares (including UK shares, stocks and shares investment savings accounts, employee shares and share options, and overseas shares): Office for National Statistics, *Wealth in Great Britain: Main results from the wealth and assets survey 2006/08* (London: Crown 2009).

172 United States SEC, *The Investor's Advocate: How the SEC protects investors, maintains market integrity, and facilitates capital formation*, available at http://www.sec.gov/about/whatwedo.shtml. Some investors may be former employees of companies who have received benefits in the form of shares: See A Hill, S McNulty and E Wine, "A poor retirement: the bankruptcy of Enron has brought into uncomfortably sharp focus broader doubts about the way that US employees invest for their old age", *Financial Times*, 12 December 2001.

practical terms, it may prove to be a token shift, given the fact that unsecured creditors often make no recoveries on their claims.[173] Furthermore, distributions aimed at compensating injured investors will be difficult to reconcile with this long-standing interpretation of the statutory *pari passu* rule:

> . . . everybody shall be paid *pari passu*, but that means everybody after the winding-up has commenced. It does not mean that the Court shall look into past transactions, and equalise all the creditors by making good to those who have not received anything a sum of money equal to that which other creditors have received. It takes them exactly as it finds them, and divides the assets amongst the creditors, paying them their dividend on their debts as they then exist.[174]

The US experience shows that the policies underlying investor protection, such as deterrence and compensation, and a regulatory body's power to enforce them, can sit uneasily with the understanding of "equality of distribution" in an insolvency context. In the UK, the *Davies Review of Issuer Liability* concluded that the question of subordinating defrauded shareholder claims needed further investigation, noting that it raised important general issues about the nature of equity investment in companies and the role of legal capital.[175] The growing correspondence between the positions of shareholders and unsecured creditors highlighted in this paper should not convey the impression that the incorporation of defrauded shareholder interests into the unsecured creditor class can be achieved through a simple grafting-on process. It carries implications for the essence of the insolvency distribution regime, and should accordingly be approached with care.

173 Noted by Gengatharen: R Gengatharen, "*Sons of Gwalia*: defrauded shareholders claims in insolvency" (2008) 17 *International Insolvency Review* 1, at p. 7. See also R J Mokal, *Corporate Insolvency Law: Theory and application* (Oxford: OUP 2005), ch. 4. The impact of parity on the cost of administering the estate may in fact further diminish their recovery levels: Hargovan and Harris, "*Sons of Gwalia*", n. 78 above, consider the implications of parity for the efficient handling of claims in insolvency.

174 *In re Smith, Knight & Co. ex parte Ashbury* (1868) LR Eq 223, at 226 (per Lord Romilly MR).

175 P Davies QC, *Davies Review of Issuer Liability: Final report* (Norwich: Stationery Office 2007), paras 61–2. It was recommended that the UK government should consider adopting the Australian CAMAC Report, n. 69 above, as part of any future policy developments in this area, but one wonders whether this recommendation would be maintained given the Australian government's advice to pursue explicit statutory subordination contrary to CAMAC's advice.

NILQ 62(1): 99–118

Not an invitation to rape: the Sexual Offences Act 2003, consent and the case of the "drunken" victim

GEORGINA FIRTH*

Law School, Lancaster University

The aim of this paper is to look at the development of the law of consent in the light of the changes in the law and consider whether the reforms have achieved their stated aims of providing "coherence and clarity". I will seek to argue that whilst there has been some progress and some recognition of the problems faced by rape complainants, changes to the substantive law alone will not make a significant difference, particularly in cases of rape involving alcohol. What is needed is a broader reform agenda which tackles problematic legal definitions but also seeks to change stereotypical perceptions of rape complainants through measures such as effective public information campaigns and the strategic use of expert evidence. In the light of these suggestions, it is argued that the decision of the government to abandon further reform proposals, such as a rebuttable presumption of non-consent where the victim is intoxicated, is a missed opportunity which has led to a return to the position prior to the Sexual Offences Act (SOA) 2003. The effect of this is that societal attitudes to rape and rape myths continue adversely to influence verdicts, to interfere with the proper consideration of the specific evidence in "difficult" cases, and to undermine the efforts of policy and legislation to bring about meaningful change.

Introduction

The law in relation to rape has received considerable attention in recent years. However, the focus of much of the reform has been on the substantive law and, in particular, the vexed issue of consent. The definition of consent set out in s. 74 of the SOA states that "a person consents if he agrees by choice and has the freedom and capacity to make that choice". The aim of enacting a statutory definition was to "bring coherence and clarity to the meaning of consent".[1] To further this aim and reinforce the new definition, in March 2006 the government launched an advertising campaign.[2] One of the advertisements contained a full-page image of a woman's torso. She is naked save for a pair of knickers with a no-entry sign emblazoned on the front. The text across the bottom of that page reads "Have sex with someone who hasn't said yes to it, and the next place you enter could be prison."[3]

* The author would like to thank the anonymous reviewer for valuable comments on an earlier draft of this paper.

1 *R* v *Bree* [2007] EWCA Crim 256, at 22.

2 www.homeoffice.gov.uk/crime-victims/reducing-crime/sexual-offences/.

3 Arguably, the text of the advertisement is also objectionable as it presents a one-dimensional way of thinking about a woman's body as a place that another enters.

Whilst the government's concern on this issue is laudable, it is somewhat disturbing that they considered that this was an appropriate image to convey the consent message. In choosing this image to get their message across, the government is, perhaps unwittingly, reinforcing the kind of stereotypes and rape myths that continue to plague the rape laws and ensure low conviction rates. The advertisement presents the woman as a body. The visual message seems to be that consent is to be read from that body and not from the context of what a woman says or her understanding of the situation (after all, she has no head with which to speak in the advert).

The problem raised by this advertising campaign – namely the ghostly silence of women's real voices and experiences from the rape trial – is one that has concerned feminists for many years[4] and it is worrying that despite substantial reform the same issues continue to arise. Recent case law, such as *Hagan*,[5] *Dougal*[6] and *Bree*,[7] indicates clearly that the problems that feminist writers have long sought to identify in relation to consent and how it should be interpreted by the law have not been resolved. This is particularly the situation in the difficult cases, for example, those where there is a degree of acquaintance between the parties or where alcohol is involved. Undeniably, it is often difficult to determine how consent is given and understood under these conditions, but when important questions of definition remain in the hands of the judiciary will the position improve? The case of *Bree* would suggest not and it is a matter of concern that this case has led to the government abandoning proposals for further revision of the consent provisions under the Act, despite considerable academic and media criticisms of judicial and juror decision-making in this area.[8]

The Home Office review and legislative change: the ghost of consent past

As a starting point, it is worth considering the review that took place prior to the enactment of the SOA and setting out what it sought to achieve in terms of the protection of particularly vulnerable victims. Although the law on rape has undergone a number of legislative and procedural changes over a period of years, it is clear from the consistently poor conviction rates that there had been little improvement from the perspective of the female rape complainant.[9] The most significant contribution to the process of reform in recent years is the SOA, which followed the Home Office consultation document *Setting the Boundaries: Reforming the law on sex offences*.[10] The two-volume document was the result of a

4 For recent feminist engagements with this subject, see R Hunter and S Cowan (eds), *Choice and Consent: Feminist engagements with law and subjectivity* (Oxford: Routledge-Cavendish 2007), chs 3, 5, 6 and 9; J Temkin and B Krahe (2008) *Sexual Assault and the Justice Gap: A question of attitude* (Oxford: Hart (2008).

5 N Britten, "Student is cleared of raping fresher he was meant to be caring for", *Daily Telegraph*, 3 November 2006.

6 *R* v *Dougal* (2005) Swansea Crown Court (unreported). In this case, the prosecution offered no evidence on the basis that a drunken consent is still consent, after the complainant testified that she could not remember whether or not she had consented to intercourse. For criticism of *Dougal*, see: C Gammell, "Judge 'dreadfully wrong' over rape ruling", *The Independent*, 24 November 2005; C Elliott and C de Than, "The case for a rational reconstruction of consent in criminal law" (2007) 70 *Modern Law Review* 225, p. 240, and the Court of Appeal in *Bree* [2007] EWCA Crim 256, para. 32.

7 [2007] EWCA Crim 256.

8 J Rozenberg, "Law on consent in rape cases is clear enough", *Daily Telegraph*, 21 June 2007. See also N S Rumney and R A Fenton, "Intoxicated consent in rape: *Bree* and juror decision-making" (2008) 71(2) *Modern Law Review* 271–302; J Temkin and A Ashworth, "The Sexual Offences Act 2003: (1) Rape, sexual assaults and the problems of consent" (2004) *Crim LR* 328–46; Elliott and de Than, "The case for a rational reconstruction", n. 6 above.

9 J Bindel, "Just another week of rapes", *The Guardian*, 29 May 2009; R Williams, "Postcode lottery in rape convictions getting worse", *The Guardian*, 10 June 2009.

10 Home Office, *Setting the Boundaries: Reforming the law on sex offences* (London: Home Office 2000).

series of consultations, deliberations and proposals arising from the Review of Sex Offences, which was established by the Home Secretary to provide recommendations for "clear and coherent offences that protect individuals . . . from abuse and exploitation, and enable abusers to be appropriately punished".[11] Whether the specific recommendations in relation to the law of rape have achieved these objectives or, indeed, whether they even adequately address the problems raised are the focus of the next section.

Both Lacey[12] and Rumney[13] criticised the remit of the government review in so far as it focused exclusively on the revision of the substantive law of rape. The consequence of this focus was inevitably that important issues relating to the way in which legal definitions of rape are interpreted and enforced by the criminal justice system and the norms, assumptions, images and values that shape that enforcement process were excluded from consideration. As Rumney has argued, the exclusionary focus serves to "obscure the practices of criminal justice agencies that hinder the effective enforcement of the law" and gives the "implicit impression" that "such problems can be resolved by [a] limited revision of the substantive law".[14]

A further and connected danger of such an approach, identified by Lacey, was that, despite the review's apparent appreciation that sexual offences can be seen as tracing the boundaries of what is socially regarded as right and wrong in the area of sexual conduct,[15] it downplayed:

> one of the most powerful aspects of criminal law . . . its reflection and reinforcement of certain ideas of the sexually normal, including the sex/gender roles and stereotypes which any informed reader of the research or observer of practice knows make a decisive difference to the implementation of sexual offences.[16]

It would therefore appear that the review's recommendations would do little to undermine the pervasive myths and stereotypes of female sexuality that have so long plagued the effectiveness of rape law from the point of view of complainants. This can be seen from both the judgment in *Bree* and the apparent lack of impact of the government's advertising campaign, referred to earlier, in changing public perceptions of "real rape".[17]

Throughout its consideration of the law of rape, the review sought to strike a balance between what it saw as "the key moral imperatives of individual autonomy",[18] bearing in mind a perceived increased liberality in social attitudes towards sex, and the "protection of the vulnerable", recognising the harmful consequences of sexual assault. Amongst the conclusions drawn as a result of this balancing exercise was a proposal to widen the scope of the definition of sexual intercourse and an attempt to uncover and then specifically define the boundaries of the constructs of consent and non-consent.[19]

---

11 Home Office, *Setting the Boundaries*, n. 10 above, Jack Straw at vol. 1, p. i.

12 N Lacey, "Beset by boundaries: the Home Office Review of Sex Offences", (2001) 3 *Crim LR*, p. 11.

13 P Rumney, "The review of sex offences and rape law reform: another false dawn?" (2001) 64 *Modern Law Review* 890, p. 890.

14 Ibid. p. 890.

15 Vol. 1, para. 1.1.3.

16 Lacey, "Beset by boundaries", n. 12 above.

17 This has since been confirmed by research into the impact of the advertising campaign. See Temkin and Krahe, *Sexual Assault*, n. 4 above.

18 Lacey, "Beset by boundaries", n. 12 above, p. 5.

19 I use the term construct here in the context of legal reform. However, I appreciate that consent is something much more than just a legal or social construct.

Consent and communicative sexuality: consent futures?

Feminist critique has long argued that the lack of a statutory definition of consent in rape has allowed gender myths and stereotypes to expand the space for male subjectivity and allowed the defendant to read off consent from the complainant's demeanour, body and sexual reputation at the expense of her voice. The review recognised and re-affirmed the centrality of consent/non-consent to the offence of rape[20] but appreciated that it was not unproblematic. It saw the problems inherent in the concept of consent as being due to the lack of a clear statutory definition. It therefore proposed to overcome these difficulties by defining consent as "free agreement"[21] and by setting out a non-exhaustive list of examples, illustrating circumstances in which consent in the sense of free agreement is not present. It was suggested that this list could form the basis of model judges' directions; for example:

> In deciding whether the complainant did freely agree to sexual intercourse . . . you should not assume that the complainant did freely agree just because they did not say or do anything . . . [or] just because they did not protest or physically resist.[22]

The review made clear that, in defining consent, it was not seeking to change its meaning[23] but to enable judges to explain clearly what the law says and juries to understand "just what is meant by consent".[24] The definition of consent approved by the review is now found in s. 74 of the SOA, which states: "A person consents if he agrees by choice and has the freedom and capacity to make that choice." This definition is supplemented by s. 75, which contains six evidential presumptions against consent, and s. 76, which contains two conclusive presumptions of non-consent.[25] Interestingly, the review's recommendation that there be an evidential presumption against consent where a person was too affected by alcohol to give free agreement[26] did not find its way into the law because of concerns by the Home Secretary that such a presumption would lead to "mischievous accusations".[27]

20 Home Office, *Setting the Boundaries*, n. 10 above, vol. 1, paras 2.10.1–3. Interestingly, the review claimed that it was not changing the *meaning* of consent but merely providing clarification. I would argue that this claim is somewhat disingenuous and that by inscribing a communicative model of sexuality into the meaning of consent, the review is indeed attempting to change the way in which consent is understood. From a feminist perspective, this is to be welcomed. However, whether this change in meaning will have a practical effect on the conduct of the rape trial in the absence of a corresponding change in the contextual, evidential and social norms is to be doubted. See also Rumney's consideration of how this shift may alter perceptions of criminal liability: Rumney, "The review", n. 13 above, p. 903.

21 Clarified by the review as a voluntary agreement between two parties to engage in sexual intercourse: Home Office, *Setting the Boundaries*, n. 10 above, vol. 1, paras 2.10.4–5.

22 Ibid. vol. 1, paras 2.11.1–6. At para. 2.10.6, the review expressed the opinion that the provision of guidelines on the meaning of consent would help both practitioners and jurors in coming to decisions in particular cases. Unfortunately, the government rejected this standard direction and the formulation of a standard direction was left to the Judicial Studies Board and the Court of Appeal.

23 Ibid. vol. 1, para. 2.10.3. See n. 33 below and the accompanying text.

24 Ibid.

25 For a critique of the provisions and, in particular, the way in which consent was framed and the working of the presumptions, see Temkin and Ashworth, "The Sexual Offences Act", n. 8 above. McEwan argues that it is not clear that those presumptions will have much effect on the process or outcome of a case in J McEwan, "Proving consent in sexual cases: legislative change and cultural evolution" (2005) 9(1) *International Journal of Evidence and Proof* 1.

26 Home Office, *Setting the Boundaries*, n. 10 above, vol. 1, para. 2.10.9.

27 Home Office, *Convicting Rapists and Protecting Victims: Justice for victims of rape* (London: Home Office 2006), ch. 3.

Communicative sexuality – a desirable but problematic notion

As Rumney notes, the use of free agreement as the basis of a definition of consent indicates the increasing influence of the concept of "communicative sexuality" within rape law.[28] Communicative sexuality gained purchase in feminist legal thought through an article in 1989 by Lois Pineau entitled "Date rape: a feminist analysis". Pineau was writing at the height of the concern in 1980s USA about the phenomenon of acquaintance or date rape.[29] Her aim was to argue for the more effective criminalisation of non-consensual sex amongst acquaintances in which there was no physical injury or threat of injury by developing a model of communicative sexuality. She critiqued the existing law on the basis that the current legal definition of consent focused on what the man thought was happening. In Pineau's view, in deciding whether or not a rape occurred, the focus should be directed towards whether or not the woman actually consented to the sexual encounter.[30] She posited that the court should therefore work backwards from actual consent to an understanding of when it is reasonable for a man to believe that there was consent.[31]

Pineau moved on from this proposition to develop criteria for consensual sex. She started from the assumption that it is not reasonable for women to agree to coerced sex or for men to believe that women have so agreed.[32] Her strategy for arguing this was, firstly, to seek to dispel the myths that posit a normal seduction as aggression and submission[33] and then, in its place, to develop a picture of a sexual relationship that undermined the current gendered view that pressured sex can also be consensual. For Pineau, the essence of a good sexual relationship is mutuality, "communicative sexual partners will not overwhelm each other with the barrage of their own desires . . . a person engaged in communicative sexuality will be most concerned with the mutuality of desire".[34]

She asserts that in sexual relationships, as in friendships, each person has a duty to respect the other and in order to develop this mutual respect and understanding there is a need for communication.[35] Aggressive or coercive sex lacks this "essence of mutuality" and, therefore, Pineau suggests that it is not sex to which it would be reasonable for a woman to consent or for a man to believe that she is consenting and thus to pursue.[36] Pineau asserts that non-consent in legal terms would therefore be made out when it can be shown that the sex involved in the allegation was not "communicative".[37] As Rumney notes,

28 Rumney, "The review", n. 13 above, p. 899.

29 L Pineau, "Date rape: a feminist analysis" (1989) 8 *Law & Philosophy* 217, p. 236. For a review of the background to the article and a consideration of Pineau's argumentation, see the introduction in L Francis, *Date Rape: Feminism, philosophy and the law* (University Park PA: Pennsylvania University Press 1996). Pineau's article is also reprinted in full in this book. Pineau's model of communicative sexuality has been the subject of a lively debate, particularly among male commentators. For examples of a variety of criticisms of Pineau's model, see chs 2–4 in Francis. For a male perspective, see M Cowling, *Date Rape and Consent* (Aldershot: Ashgate 1998).

30 Pineau, "Date Rape", n. 29 above, p. 224.

31 Ibid. pp. 234–5.

32 Ibid.

33 Ibid. pp. 225–33.

34 Ibid. p. 236.

35 Ibid. p. 235.

36 Ibid. p. 234.

37 Ibid. pp. 239–40. Critics of Pineau have sought to argue that this model of sexuality fails to encompass women who enjoy submissive sex. However, I would argue that there is nothing in Pineau's model that prevents these women communicating this desire to their partners in advance. For an example of Pineau's model in action, see the Antioch College Sexual Offence Policy discussed in Francis, *Date Rape*, n. 29 above, pp. xviii, 65–7, 135–75.

at face value, the notion of free agreement may have benefits for rape complainants seeking to prove their cases in that it challenges the assumption that a woman is available for intercourse when she is in no position to choose and the belief that coercion may be an acceptable means by which to force a woman into sex.[38] This could also have repercussions for cases where women are so heavily intoxicated that it is arguable that their ability to make choices is significantly impaired. However, that will depend upon a robust interpretation of capacity to consent as I will discuss later.

Feminist legal thought has long argued that free agreement involving notions such as respect for the autonomy of others and sexual integrity should lie at the heart of any legal definition of rape and its enforcement.[39] Free agreement clearly has the potential to be more effective in constructing sexual encounters where there is no agreement, verbal or otherwise, as rape.[40] This may explain Rumney's observation that the review's recommendations concerning free agreement appear to have received little critical comment.[41] Many other jurisdictions had already adopted the notion of communicative sexuality in their definitions of consent[42] and the review had access to information about the content and structure of this legislation in a number of jurisdictions.[43] However, the available research on the operation of the model in other jurisdictions would seem to suggest that the inscription of communicative sexuality into the core concepts of rape law is not unproblematic, both ideologically and procedurally.[44]

In her original article and in a response to her critics,[45] Pineau suggested that her model of communicative sexuality would open the way for rape prosecutors to shift the focus of blame from the complainant to the defendant through questioning that focused on what actions the accused took to ensure that she was consenting.[46] The government review seemed to have taken Pineau's argument on board when it stated that such a shift would take place as a result of their definition of consent.[47] However, I would argue that the current reality of rape trials in England is such that the claim that the focus of rape trials will shift because of this new definition alone is problematic. The government initially appeared to recognise this in the recent consultation paper *Convicting Rapists and Protecting Victims – Justice for victims of rape*.[48] However, the fact that it has now abandoned further proposals for reform of the substantive law following *Bree* may be a missed opportunity.

38 Rumney, "The review", n. 13 above, p. 899.

39 N Lacey, *Unspeakable Subjects: Feminist essays in legal and social theory* (Oxford: Hart 1998); N Naffine, "Possession: erotic love in the law of rape" (1994) *Modern Law Review* 10.

40 Some commentators have argued that free agreement implies something more than silence and that it may be problematic where a woman remains silent, not through fear but through a willingness to engage in sex. However, Rumney suggests, and I would concur, that consensual yet silent sex is unlikely to give rise to an allegation of rape. Rumney, "The review", n. 13 above, p. 903.

41 Ibid. p. 899. Rumney here is referring specifically to Lacey, "Beset by boundaries", n. 12 above; M Bowley, "New Labour, new sex", *NLJ*, 28 July 2000, p. 1134; J Temkin, "Getting it right, sexual offences law reform", *NLJ*, 4 August 2000, p. 1169.

42 Rumney, "The review", n. 13 above, pp. 899–901.

43 Home Office, *Setting the Boundaries*, n. 10 above, vol. 2, Appendices E, F and G.

44 For a critique of the Australian and New South Wales models, Naffine, "Possession", n. 39 above, pp. 23–9. In relation to the US model, see P R Sanday, *A Woman Scorned: Acquaintance rape on trial* (London: University of California Press 1996), ch. 12.

45 L Pineau, "A response to my critics", in Francis, *Date Rape*, n. 29 above, ch. 5.

46 Pineau, "Date rape", n. 29 above, pp. 240–1.

47 Home Office, *Setting the Boundaries*, n. 10 above, vol. 1, para. 2.11.4.

48 Home Office, *Convicting Rapists*, n. 27 above.

To assume that a cultural and procedural change in the trial process would occur overnight through the enactment of a statutory definition alone would be somewhat naive. Rumney suggests, and this has to some extent been borne out, that the communicative model may be helpful in cases where the complainant suffers serious physical harm, where there is corroborative evidence, or where the woman is incapable of communicating agreement because she is unconscious through, for example, drugs or drink.[49] However, as he notes, the court very often interprets these scenarios as "real rapes" in any event and, thus, convictions are less difficult to obtain.[50]

In the area of acquaintance rape, where the law has great difficulty in distinguishing between consensual sex and rape, it is difficult to envisage how this new definition would shift the focus decisively away from the complainant without further elaboration. She still has to give evidence and be cross-examined. Her behaviour, dress and demeanour are still subject to detailed examination and comment by the defence counsel "in order to establish whether or not there was agreement, the nature and content of that agreement and whether there was any ambiguity in what the complainant said or did".[51] Further, even though the SOA departed from the subjective *Morgan* defence of honest belief[52] and replaced it with a test of reasonable belief, this belief has been interpreted as partly subjective and therefore the defendant's point of view as to whether or not there was free agreement remains the legal point of view and, thus, possibly decisive when this defence is run.[53] This was confirmed by the Court of Appeal in the recent decision of *Bree*.[54]

Ideologically, the notion of free agreement clearly raises the issue of the "respective responsibilities of men and women".[55] Several commentators have suggested that the standard will not only require men to ensure that they have consent but also place a burden on women to communicate clearly. This is recognised by the review.[56] However, this implicit requirement may be potentially problematic as it may underline beliefs that women are in some way responsible for rape when they have not clearly and perhaps verbally and physically, through resistance, expressed their non-consent. As Rumney notes, "given the prevalence of attitudes within both society and the criminal justice system which attribute

49 Rumney, "The review", n. 13 above, p. 900.

50 Ibid. p. 900.

51 Ibid. p. 901. Carol Smart, in *Feminism and the Power of Law* (London: Routledge 1989), pp. 34 and 42, notes that telling the story of rape inevitably reveals ambiguities, for example, a woman may agree to intimacy but not intercourse. In a trial when credibility is in issue, these ambiguities will be used to discredit her account of rape and subsume the rape into the "normal" pattern of submission and the individual woman into a single category of "Woman" known to be "capricious and mendacious".

52 Lacey notes that the review justified its stance by asserting the disproportionate costs to the defendant and victim of establishing consent on the one hand and of a mistake being made on the other: Lacey, "Beset by boundaries", n. 12 above, p. 12.

53 Whilst the review suggested that the controversial *Morgan* defence should remain, it did recommend that the defence should not be available if the accused did not take all reasonable steps in the circumstances to ascertain free agreement at the time; or where the defendant was in a state of self-induced intoxication, or if the defendant was reckless: Home Office, *Setting the Boundaries*, n. 10 above, vol. 1 para. 2.13.8. Contrary to the review's recommendations, the Act replaced the *Morgan* test with a test of reasonable belief. However, the courts have applied the test of reasonableness in a similar way to the reasonableness requirements of the defence of provocation. This means that characteristics of the defendant will be taken into account. See Temkin and Ashworth, "The Sexual Offences Act", n. 8 above.

54 *R* v *Bree* [2007] 2 All ER 676, para. 34. On this point see, J Elvin, "Intoxication, capacity to consent and the Sexual Offences Act 2003" (2007) *King's Law Journal* 151, p. 156.

55 Rumney, "The review", n. 13 above, p. 901, and Pineau, "Date rape", n. 29 above, pp. 234–5.

56 Home Office, *Setting the Boundaries*, n. 10 above, vol. 1, para. 2.10.3.

blame for sexual violence to women[57] (it is not unreasonable to assume that) any notion of responsibility may disproportionately fall on the woman".[58] I will return to this issue of responsibility later in relation to cases of rape involving alcohol.

A further potential difficulty, inherent in the notion of free agreement, is that it appears to place considerable emphasis on the nature and quality of communication. Research evidence from Canada and the United States[59] has clearly indicated that current gendered notions of the normality of a coercive model of sexual intercourse mean that men (and women) have internalised gendered myths about women's willingness to engage in sexual intercourse. For example, that a woman who initiates a date, enters a man's flat or accepts dinner from him is expressing a willingness to sleep with him or even that such actions entitle a man to sleep with a woman against her wishes.[60] The review potentially added to these problems by recognising both verbal and non-verbal sexual consent.[61] This wide notion of communicative sexuality is arguably positive in that it recognises the reality of sexual relations.[62] However, it also potentially allows defendants to contradict complainants' verbal expressions of non-consent by feeding into the myths and stereotypes already referred to. Such myths and stereotypes continue to play an important if hidden role in the trial process and, thus, the notion of free agreement alone is unlikely to unequivocally benefit complainants or substantially increase the protection offered by the criminal law.[63]

Despite the inferences that can be drawn following the changes in the law in the Criminal Justice and Public Order Act 1994, s. 34, there is still no requirement that the defendant give evidence in English law, even though the new definition requires in s. 1(2) that the jury consider the defendant's reasonable belief in "all of the circumstances including any steps the defendant has taken to ascertain whether the victim consents".[64] Furthermore, there is no guarantee that a complainant's account will withstand cross-

57 See M Amir, *Patterns in Forcible Rape* (London: University of Chicago Press 1971) in relation to notions of complainants' contributory negligence.

58 Rumney, "The review", n. 13 above, p. 901.

59 For example, see: T P Humphreys and E Herold, "Date rape: a comparative analysis and integration of theory" (summer 1996) 5(2) *Canadian Journal of Human Sexuality* 69; D M Truman, D M Tokar and A R Fischer, "Dimensions of masculinity: relations to date rape supportive attitudes and sexual aggression in dating situations" (July–August 1996) 74 *Journal of Counselling & Development* 555; V Van Wie, A M Gross and B P Marx, "Females' perception of date rape: an examination of two contextual variables" (December 1995) 1(4) *Violence Against Women* 351.

60 Ibid.

61 Home Office, *Setting the Boundaries*, n. 10 above, vol. 1, para. 2.10.4. Some authors have argued that nothing less than a specific "yes" should count as consent. See S Schulhofer, *Unwanted Sex: The culture of intimidation and the failure of law* (London: Harvard University Press 1998). However, other commentators have argued that communicative sexuality as a concept is objectionable because it appears to focus less on the presence of consent than on how someone should consent to sex: see H Reece, "When a woman says 'no' she means 'no'", *NLJ*, 7 November 1997, p. 1616. I would, however, agree with Rumney and the Campaign against Rape, when they argue that Reece has misunderstood communicative sexuality. There is nothing in Pineau's work to suggest that she wishes to stipulate where people should have sex or what sexual positions they should adopt. For Pineau, these are matters of individual choice based upon mutual communication and respect.

62 For a full discussion, see D Archard, *Sexual Consent* (Boulder: Westview Press 1998).

63 J Kelly, J Lovett and L Regan, *A Gap or a Chasm? Attrition in reported rape cases*, Home Office Research Study 293 (London: Home Office 2005); E Finch and V E Munro, "Juror stereotypes and blame attribution in rape cases involving intoxication: the findings of a pilot study" (2004) 44 *BJ Crim* 1–14; E Finch and V E Munro, "The Sexual Offences Act 2003: intoxicated consent and drug assisted rape revisited" (2004) *Crim LR* 789; E Finch and V E Munro, "Breaking boundaries? Sexual consent in the jury room" (2006) 26(3) *Legal Studies* 303–20.

64 Criminal Justice and Public Order Act 1994, s. 34; Sexual Offences Act 2003, s. 1(2).

examination[65] or that there will be sufficient corroborative evidence to persuade a jury to convict.[66] Therefore, the evidential and procedural standards that continue to govern the trial process will ensure that the primary focus of rape trials will continue to be on the complainant and the defendant's perception of her state of mind. Given the current high attrition rates and the attitudes of the police and Crown Prosecution Service to rape allegations, any refocusing of the trial process is unlikely to benefit many complainants without a concomitant refocusing of societal and institutional attitudes to acquaintance rape, as will become clear when the case of *Bree* is considered below.[67]

The case of the intoxicated victim: consent unravelled again or the spectre of the past?

Recent cases[68] decided using the new definition of consent appear to demonstrate that the concerns raised at the time of the reform have now been realised and that, in isolation, the new definition of consent has not had a significant effect on conviction rates in rape cases.[69] Arguably, many of the cases that have been reported in the media involve particularly difficult issues, combining the question of acquaintance (however brief) and the role of alcohol in sexual encounters. However, it is precisely these difficult cases which have contributed to the high attrition rates in rape cases and prompted the need for reform in this area, so it would seem that they are an important indicator of the success or otherwise of the new definition.[70] Several of these cases (*Dougal* in particular) raised renewed concern about the effectiveness of the definition of consent in the SOA and in particular, whether it can offer protection to women in especially vulnerable situations. These concerns were reinforced by research that indicates that juries are reluctant to convict men of rape in cases in which the victim has been drinking.[71]

As a result of the publicity surrounding cases such as *Dougal*, the Home Office revisited the issue of consent in the consultation paper, *Convicting Rapists and Protecting Victims: Justice for*

65 Although there does appear to be evidence that false allegations of rape are rare. See S Lees, *Carnal Knowledge: Rape on trial* (London: Penguin 1996).

66 Rumney, "The review", n. 13 above, p. 900. Z Adler, *Rape on Trial* (London: Routledge & Kegan Paul 1987); Lees, *Carnal Knowledge*, n. 65 above; and J Temkin, "Prosecuting and defending rape: perspectives from the Bar" (2000) 27 *Journal of Law and Society* 219 have all recorded the marked reluctance of juries to convict defendants in cases involving a degree of acquaintance. This, of course, corresponds with the widespread societal views shown in psychological studies that such assaults are less serious than assaults by strangers and, indeed, in some circumstances, that they are not rape at all.

67 Rumney, "The review", n. 13 above, fully addresses the limitations of the review in relation to the procedural and evidential difficulties that were not addressed, for example, victim unwillingness to seek legal redress, attitudes and practices of criminal justice personnel and jurors' attitudes (pp. 904–9). These issues have already been addressed to some extent elsewhere and therefore it did not seem appropriate or necessary to re-address them here. Rumney concludes, and I would agree, "*effective legal change will only occur when there is an inclusive process of law reform that examines formal legal rules, their interpretation and enforcement*" (p. 910). See also Lacey, "Beset by boundaries", n. 12 above, pp. 11–13.

68 *R* v *Dougal* (2005) Swansea Crown Court (unreported); *R* v *Hagan* (2006) Nottingham Crown Court (unreported). For media reports of these cases see: G Roberts, "Drunken consent to sex is still consent, judge rules", *The Independent*, 24 November 2005; N Britten, "Student is cleared", n. 5 above.

69 R Williams, "Postcode lottery", n. 9 above.

70 In relation to attrition rates generally, see L Kelly, *Rape in the 21st Century: Old behaviours, new contexts and emerging patterns*, Full Research Report, ESRC End of Award Report, RES-000-22-1679 (Swindon: ESRC 2007).

71 Finch and Munro, "Breaking boundaries?", n. 63 above, p. 306; Finch and Munro, "The Sexual Offences Act 2003", n. 63 above. The research indicated that juries believed that it was reasonable for a man to assume that a woman's silence amounted to consent, even when it was due to intoxication. They tended to hold a drunken victim partially responsible for what happened and they were less inclined to see taking advantage of a drunken woman as rape.

victims of rape.[72] One of the key proposals in the paper was to consider whether consent needs further clarification in relation to the question of capacity, either in the form of a further definition and/or the creation of an evidential presumption.[73] However, the judiciary expressed considerable hostility to the proposals on consent in the consultation paper and the Court of Appeal took the opportunity in the case of *Bree*[74] to undermine the consultation process and reclaim questions of interpretation as matters best left to the courts.[75]

Benjamin Bree had been convicted of the rape of a 19-year-old student. In brief, the facts of the case were that he had spent an evening drinking with the student, who shared a flat with his brother. On their return to the flat, the complainant was sick and Bree helped her to wash her hair. Her next memory was of finding herself on the bed with Bree having sex with her. She told the jury that she had not consented. Bree's defence was that the sex was consensual and that he reasonably believed that she was consenting. The Court of Appeal held that the essential question in this case was not whether the parties to the case were less inhibited because they had been drinking, no whether they might have regretted what had happened, nor whether they had acted irresponsibly, but simply whether the evidence proved that the appellant had intercourse with the complainant without her consent.[76] The judgment states that neither party had acted unlawfully in drinking to excess and that both were free, if they wished, to have intercourse with each other. The court qualified the position somewhat saying that if, through drink or any other reason, the complainant has temporarily lost her capacity to choose whether to have intercourse on the particular occasion, she is not consenting and, subject to questions about the defendant's state of mind, if intercourse took place, this would be rape. However, where the complainant has voluntarily consumed even substantial quantities of alcohol but nevertheless remains capable of choosing whether or not to have intercourse and in drink agrees to do so, this would not be rape.[77] Finally, the court concluded that there was a clear definition of consent and that to try to set a level of intoxication that would negate consent would not be possible. Otherwise, "provisions intended to protect women from sexual assaults might very well be conflated into a system, which would provide patronising interference with the rights of autonomous adults to make personal decisions for themselves".[78]

The judgment appears, on the face of it, to be attempting to protect individualistic notions of personal and sexual autonomy. However, I would argue that the Court of Appeal in *Bree* dismissed the numerous concerns[79] that had been raised about the new definition too easily, situating the difficulty as one raised by the "infinite circumstances of human

72 Home Office, *Convicting Rapists*, n. 27 above.

73 Home Office, *Convicting Rapists*, n. 27 above, ch. 3. As noted above, this was not a new concern and the review prior to the SOA 2003 had recommended that a person was unable to consent when too affected by alcohol to give free agreement (para. 2.10.9). This suggestion was not included as a presumption under s. 75 due to a concern expressed by the Home Secretary about false accusations.

74 *R* v *Bree* [2007] 2 All ER 676.

75 J Rozenberg, "Judges reject government rape reforms", *The Telegraph*, 24 January 2007; C Dyer, "Judges try to block rape trial reforms", *The Guardian*, 23 January 2007.

76 *R* v *Bree* [2007] 2 All ER 676, para. 26.

77 Ibid. para. 34. The guidelines set out in *Bree* can now be found in the specimen direction in cases involving intoxication in the March 2010 version of the *Crown Court Bench Book* at www.jsbni.com/Publications/BenchBookRevisedApr2010.doc.

78 Ibid.

79 For example, Temkin and Ashworth, "The Sexual Offences Act", n. 8 above; Elliott and de Than, "The case for a rational reconstruction", n. 6 above; Finch and Munro, "The Sexual Offences Act 2003", n. 63 above.

behaviour"[80] rather than lack of clear statutory guidance.[81] The judgment can be seen as problematic in two ways. Firstly, whilst there are some attempts to focus on the legal issues at stake,[82] it fails to take proper account of the fact that the legal rules are not always applied and interpreted by juries free of entrenched attitudes and behaviours which may be damaging to rape complainants in cases involving alcohol. Secondly, the court failed to take the opportunity to clarify the issue of the definition of capacity and to map this on to a legal definition of consent in a meaningful way.

Capacity and consent: when should drunken consent equate to legal consent?

The main issue at appeal was the lack of guidance on the issue of consent from the trial judge to the jury and the judgment of the Court of Appeal agreed that the jury should be given assistance on the question of capacity and the extent it can take this into account when determining the issue of consent. However, it is noteworthy that nowhere in the judgment did the court set out to clearly define capacity or to provide a model direction on consent for trial judges to follow so it remains unclear what guidance on these issues jurors should be given in future cases. The specimen direction on capacity in the 2010 edition of the *Crown Court Bench Book* follows *Bree* on this, in that it simply states that "where a question of capacity arises in the evidence it must be left for the jury to decide".[83] In the course of the judgment, the Court of Appeal approved the pre-2003 case law and I suggest it is likely that in seeking to guide juries in the future, trial judges will resort to this case law and the pre-2003 definitions of consent that have been problematic for complainants as too much discretion is left to jurors and their interpretation of what "satisfies the ordinary meaning of consent".[84]

The judgment in *Bree* starts by setting out the relevant provisions of the 2003 Act and deals with some of the debate prior to the Act. In this way, the Court of Appeal is indicating that it is aware of the debates and is sensitive to the issues that led to reform. However, the judgment immediately seeks to legitimate and reclaim the interpretative role of the judiciary in relation to the legislation by situating the difficulty in these cases as one of proof and not ideology and, in doing so, it sidesteps the need to further define the capacity element of the new definition. It notes that the Act is intentionally silent on the question of consent and the voluntary consumption of alcohol and suggests that this is because the law was already clear.[85] Sir Igor Judge states that the presumptions negating consent set up in ss. 75 and 76 of the Act deal with matters of "common sense". It is held to be obvious that there would be no consent where the victim is insensible and equally, it is a matter of common sense that voluntary intoxication was not covered by the legislation, as it is a complex matter of proof to be decided by reference to the facts on a case-by-case basis. In this way, the judgment is using common sense and questions of proof as rhetorical devices to reclaim the judicial high ground, avoid the need for further legislative intervention (such as a presumption of non-consent when the victim is heavily intoxicated, or a more generous

80 *R* v *Bree* [2007] 2 All ER 676, para. 36.

81 Rumney and Fenton, "Intoxicated consent", n. 8 above.

82 *R* v *Bree* [2007] 2All ER 676, paras 25–6.

83 *Crown Court Bench Book*, n. 77 above, p. 371. See also *Hysa* [2007] EWCA Crim 2056.

84 Elliott and de Than, "The case for a rational reconstruction", n. 6 above, p. 234. See also Rumney and Fenton, "Intoxicated consent", n. 8 above; Elvin, "Intoxication", n. 54 above, pp. 151–7.

85 Para. 27 of the judgment is important in this respect. The Court of Appeal states that before the 2003 Act, it was not difficult to identify the relevant legal principle and for the judge to explain the law. It approves judicial directions in *Malone* [1998] 2 CAR 447 and *Lang* [1976] 62 CAR 50 and disapproves at the same time any question of further legal definition imposed by government. The clear implication is that there is no need for any further definition as the judiciary filled the gap in the law 30 years ago.

interpretation of capacity) and state that the judges are best placed to determine where the line between consent and non-consent should be drawn.

In relation to intoxicated complainants who are less than insensible, the judgment explicitly rejects the Home Office's argument that consent or capacity has a special meaning that should be determined by legislation and argues that such cases are best decided by relying on the ordinary meaning of the term consent. Sir Igor Judge notes that questions about consent when affected by voluntary intoxication can cover a wide range of states of mind and as such are social matters rather than something that lends itself to legislation.[86] It can be seen that, without explicitly stating that it is doing so, by linking the term consent with a state of mind that should be determined on the facts of the individual case, the Court of Appeal is seeking to import into s. 74 the definition of consent found in *Olugboja*.[87] The consequence of this is that, as in *Olugboja*, the judiciary is placing itself in a position of authority. There are clear indications throughout the judgment in *Bree* that determining the correct level at which intoxicated consent is no longer "real" consent is not something that the jury can do unaided, after all, that is what happened in the Crown Court in this case and led to the appeal.[88] It is a matter for the judge and the jury is to accept the judge's analysis of consent up to that point. The Court of Appeal appears to be arguing that setting the appropriate level of voluntary intoxication that negates consent would be too complex a matter for legislation because of the infinite variety in the facts of a particular case. For the Court of Appeal, the ordinary meaning of consent is a matter of common sense, is sufficiently clear, should depend on the individual facts of the case and is a matter for the jury to fix by applying its good sense and knowledge of human behaviour to the facts of the case.

The question that therefore arises is: what is the ordinary meaning of consent that the judiciary is applying here? Is it the complainant's state of mind that is in issue or could it be the defendant's state of mind, taking into account what he reasonably believed? How are the jury to determine when the line between real but intoxicated consent and rape is crossed? The Court of Appeal refused to fix a formula. It argued that it is a matter for the jury taking into account standards with which the jury is said to be familiar.[89] However, to prevent the jury fixing the boundary in the wrong place, it is stated that the role of the judge is to steer the jury to the right point through the discourses in the summing up.[90] In reclaiming this role for itself, the judiciary is inferring that its understanding of common sense is drawn from and encompasses societal and cultural notions of what is right and appropriate. Common sense is a rhetorical device used to indicate the difference between real rape where the victim is unconscious (and there is little difficulty in proving the fact of rape) and the situation where the victim has voluntarily become intoxicated, consent is in issue and proof is more sensitive and difficult. In reclaiming an interpretative role, the judge is placed in a position of authority as the most likely person to be able to find the truth of an allegation. However, as Wallerstein notes, "in the case of the capability of the drunken

86 In making this argument, the Court of Appeal rather disingenuously relies on the article by Temkin and Ashworth, which notes the philosophical difficulty in defining concepts such as freedom, choice and capacity and points out that the definitions remain "at large": Temkin and Ashworth, "The Sexual Offences Act", n. 8 above, p. 336.

87 Gardner's analysis of *Olugboja* clearly makes this link between consent and state of mind: S Gardner, "Appreciating *Olugboja*" (1996) 16(3) *Legal Studies* 275.

88 *R* v *Bree* [2007] 2 All ER 676, paras 39–42.

89 Ibid. paras 35 and 36.

90 Ibid. para. 39.

woman to consent to sex, the plummeting rate of rape convictions proves that common sense seems to fail".[91]

In seeking to address the criticism that arose following the case of *Dougal*, and in order to justify its position, the court built on the analogy between drunken consent and drunken intent that had been raised in the earlier case and which led to the phrase "a drunken consent is still consent". However, as Wallerstein argues, such an analogy, which equates intoxicated intent with consent in the criminal law, is at the very least controversial and potentially misleading.[92] She suggests that this is because the rationale underlying the law on intoxication applies to acts done by an *offender* that lead to harm to another. The issue at stake, however, in cases such as *Bree* are the actions of voluntarily intoxicated *victims*.

> To equate the victim's consent to the offender's intent is to extend the rationale of the law relating to the offender's intoxication onto the drunken victim [and] to imply that there can be no recognition of the victim's drunkenness as negating her consent because there is some fault in voluntarily getting drunk.[93]

Arguably, a woman might put herself in a vulnerable position if she chooses to get drunk but this does not mean that she is making an indirect choice to have sex. In making this link between the terms intoxicated intent and consent, there are suggestions in the judgment that when a woman chooses to get drunk, she is implicitly choosing to have sex. This reflects a worrying set of assumptions and values as I shall discuss further below.[94]

In contrast to the opinion expressed by the Court of Appeal and some commentators,[95] I would suggest that one of the key rationales behind the judgment, the statement that "drunken consent is still consent", is too simplistic and is problematic for several reasons. Whilst *Bree* established that intoxication does not necessarily deprive a person of the capacity to consent, it does not attempt to give guidance on the point at which capacity to consent is lost.[96] As was made clear by Sir Igor Judge, the presumptions in ss. 75 and 76 will not apply in these difficult cases and so the s. 74 definition is central. I would not dispute that identifying the point at which a complainant becomes incapable of consent does depend in part on the facts and the available evidence.[97] However, I would suggest the definition in s. 74 requires more than this. Even when a person appears to have agreed as a matter of fact on the available evidence, what needs to be considered is whether they are capable of agreeing and this necessitates a *legal* interpretation of the key issue of capacity in the case of a conscious but very drunk victim. This is because the definition of consent set out in s. 74 is clearly intended to require more than factual agreement as it includes the terms "capacity" and "freedom to consent". It is only when all three conditions are met that consent in law is present. As Wallerstein argues:

> The question of whether a person is capable of agreeing must therefore be understood as a legal question about the conditions of mental capacity under which the law would recognise factual consent as a basis of legal consent.[98]

91 S Wallerstein, "A drunken consent is still consent – or is it? A critical analysis of the law on a drunken consent to sex following *Bree*" (2009) 73 *JCL* 318–44, p. 343.

92 S Wallerstein, "A drunken consent", n. 91 above, p. 323.

93 Ibid. p. 326.

94 For a fuller consideration of the arguments against conflating intent and consent in this way see ibid.

95 See, for example, J Elvin, "The concept of consent under the Sexual Offences Act 2003" (2008) 72 *JCL* 519.

96 *R* v *Bree* [2007] 2 All ER 676, para. 39.

97 Rumney and Fenton, "Intoxicated consent", n. 8 above, p. 283.

98 Wallerstein, "Drunken consent", n. 91 above, p. 321.

In *Bree*, whilst the court recognised in principle the "vanishing capacity to consent",[99] it did not go on to develop this point or apply it to the facts of the case.[100] Instead, the focus of the judgment in relation to the voluntarily intoxicated victim in this case was interpreted restrictively to be a question of a lack of "actual" consent rather than lack of capacity to consent.[101] In this way, the court was able to focus on issues of proof and avoid the question of the need for further definition. Given that the victim in this case was, as the judgment recognised, extremely drunk, it would therefore seem post-*Bree*, that the law requires proof of a very high degree of intoxication, almost to the point of unconsciousness, before capacity to consent is negated. Does this really reflect the fair balance between recognising sexual autonomy and protecting the vulnerable that the review was attempting to achieve with the new definition, or should a much lower level of intoxication be required to negate capacity to consent? Alternatively, given the difficulties in establishing the point at which capacity is lost, should there be a default position set out in a presumption that drunken consent is not consent?[102] This might be one way of tackling the hidden assumptions that influence so many judgments in cases of rape involving alcohol and which informed the decision in *Bree* as I will now move on to discuss.

Rape myths and alcohol

In taking pains to be seen not to patronise women, it could be argued that the court placed too much emphasis on what might be termed "positive autonomy" at the expense of recognising "negative autonomy",[103] or the right to say no and have your no taken seriously. This latter emphasis may have been a consequence of the judicial hostility to further legislative intervention but as, Finch and Munro's studies have shown,[104] lack of clear legal principles and guidance on them can lead to juries relying on rape myths to inform their understanding and, thus, further undermine women's claims of non-consent in particular cases.

The issue of sexual autonomy is a complex one and there are undoubtedly tensions between the recognition of positive autonomy and the protection of negative autonomy. However, it could be argued that the court's focus on positive autonomy sends out the wrong message and fails to recognise the vulnerability of victims whose choices may be impaired by their alcohol intake.[105] The argument that setting high standards for what qualifies as consent would deprive women of their autonomy is only valid if we ignore the powerful role of the law in setting standards and assume that sexual autonomy in all of its

99 Wallerstein, "Drunken consent", n. 91 above, p. 321.

100 However, see the later case of *R* v *H* [2007] EWCA Crim 2056 where the court did state that capacity to consent is a legal question independent of the question of unconsciousness.

101 It was noted that during the trial, the prosecution had, in the light of the complainant's evidence, changed its position from suggesting that the complainant was so drunk that she was unconscious to the position that she was conscious, had only a patchy memory of events, but nevertheless did not consent. The Court of Appeal in *Bree* suggested that this change was a move away from claims of lack of capacity to a claim of lack of actual consent. This analysis is questionable.

102 Wallerstein argues for a default position that drunken consent is not consent. However, she seems to propose that this should form part of the legal definition of capacity in s. 74 rather than a separate presumption of non-consent: Wallerstein, "Drunken consent", n. 91 above.

103 Elliott and de Than, "The case for a rational reconstruction", n. 6 above, p. 231. For a further discussion of this issue see A Wertheimer, *Consent to Sexual Relations* (Cambridge: CUP 2003); and Elvin, "Intoxication", n. 54 above, pp. 151–7.

104 Finch and Munro, "Juror stereotypes", "The Sexual Offences Act", and "Breaking boundaries", n. 63 above. See also Rumney and Fenton, "Intoxicated consent", n. 8 above, p. 289.

105 Throughout the judgment, stress is laid on the complainant's voluntary consumption of alcohol. For example, see paras 25, 26, 39. This was contrasted with the presumptions which deal with complainants who are *involuntarily* at an advantage (para. 24).

aspects is a right that can be straightforwardly exercised. However, given the connections drawn between women's drinking and rape in the media,[106] the political discourses and even in the government's own campaigns,[107] it would seem that women's "right" to drink and to consent to sex or not when they have been drinking is far from straightforward.

It is clear from the judgment that the court implicitly accepts that alcohol is a normal element of socio-sexual relations and it is at pains to make clear that it is not suggesting that a person cannot consent to sex when they have been drinking.[108] However, in accepting the normalisation of the connection between alcohol and sexual relations, the court fails to address the damage done by the myths about rape involving alcohol which have arguably been at the root of the problem with conviction rates. A growing body of research has recognised that rape convictions are particularly low in cases where alcohol is involved due to the fact that juries are influenced by cultural discourses which can lead to stereotypical views of appropriate gender behaviour.[109] Examples of this are the myths that women lack the ability to directly communicate sexual desire and that they become more sexually promiscuous when drunk. Numerous studies have shown that women who become part of a drinking culture, or who flirt when drunk, are considered "loose" by a proportion of the population.[110] From this, as Meyer has shown in relation to an analysis of a number of media articles, it is a small step to equate consenting to get drunk in the company of men to consenting to sex and to holding women accountable for rapes involving alcohol.[111] Thus, rape involving alcohol becomes re-presented as a problem of female drinking rather than male violence. As Meyer has argued, female drinking then becomes integrated into the blame culture because it is argued that it impairs the woman's ability to provide for her own safety, makes her vulnerable and, thus, constitutes irresponsible and unwise behaviour.[112] The language of the judgment in confirming and supporting the phrase "drunken consent is still consent" seems almost to be legitimising the practice of drunken sex and setting it up as a question of mutual responsibility.[113] This notion of mutual responsibility is to some extent, as discussed earlier, consistent with the notions of free agreement which are behind the definition in s. 74. However, in taking this view, the Court of Appeal (and the review that drafted the new definition) fails to clearly address the fact that the risks of heavy drinking mostly accrue to women, as juries may interpret the behaviour of women who drink heavily in the company of men as "asking for it". Stating that both parties were autonomous adults[114] fails to recognise this fact and risks undermining aspects of the very autonomy that the court claims it is at pains to protect.

Further, contrary to research showing that rape in an acquaintance situation is more harmful in some ways than stranger rapes, because of the breach of trust involved,[115] such myths reinforce the idea that rape involving alcohol is not serious because the victim's

106 A Meyer, "Too drunk to say no" (2010) 10(1) *Feminist Media Studies* 19–34.

107 See A Wollenberg, "Blaming the victim", *The Guardian*, 8 July 2008, on the advice given out in an advert from Directgov on drinking and rape, which seems to imply that if you are drunk and a victim of rape then you are to blame.

108 *R* v *Bree* [2007] 2 All ER 676, para. 25.

109 For example, see Kelly et al., *A Gap*, n. 63 above; Finch and Munro, "Juror stereotypes", "The Sexual Offences Act" and "Breaking boundaries?", n. 63 above.

110 Kelly et al., *A Gap*, n. 63 above; Finch and Munro, "Juror stereotypes", "The Sexual Offences Act" and "Breaking boundaries?", n. 63 above.

111 Meyer, "Too drunk", n. 106 above, at p. 26.

112 Ibid. p. 29.

113 *R* v *Bree* [2007] 2 All ER 676, paras 25 and 26.

114 Ibid.

115 Home Office, *Setting the Boundaries*, n. 10 above.

drinking was voluntary.[116] This voluntariness is something that is stressed by the Court of Appeal throughout the judgment and it makes a clear distinction between the complainant's voluntary intoxication in this case and the presumptions which the court states are designed to deal with victims who are involuntarily at a disadvantage.

The judgment fails to appreciate, when allocating responsibility for risky behaviour, that men do not bear an equivalent risk and they are not expected to modify their behaviour.[117] This position therefore questions women's victim status, as they have failed to protect themselves. It also fails to recognise that it is not simply the consumption of alcohol that produces vulnerability, but also the setting in which it occurs, namely men's reactions to women drinking.[118] In this way, the court implies that rape involving alcohol becomes a problem of female heavy drinking and not male rape. Women are seen as accountable not only for their own behaviour in making themselves vulnerable but also for not being able to stop any potential rapist.[119] Such reasoning ignores the fact that when women (victims) are at risk, the man (offender) is the risk. It might be argued that shared responsibility in these situations is only fair, but I would suggest that in placing the behaviour of the victim and offender on an apparently equal footing, the court is drawing a "flawed comparison of vastly different situations as if they had the same meaning and consequences".[120] This problem is compounded in the courtroom situation as the right to get drunk and consent to sex turns into the victim's responsibility to prove her non-consent and justify her actions because the burden of proof rests on her and an assessment of her evidence.[121] This was seen in *Dougal*, where evidence of impaired memory due to heavy intoxication on the part of the victim, rather than being seen as evidence to support her lack of capacity, was fatal to the prosecution.[122]

Reconsidering reform – or drinking is not a crime: rape is

Bearing in mind the problems with the judgment in *Bree* set out above, it is suggested that the addition to the legislation of a rebuttable presumption that drunken consent is not consent in cases of heavy voluntary intoxication should be reconsidered. As was argued above, the end result of voluntary intoxication may put a complainant in a disadvantageous situation similar to those already covered by the existing presumptions under s. 75. Perhaps more importantly, the reasons given for omitting to include such a presumption – in particular the risk of false allegations being made, are not supported by the available research evidence.[123] Neither are concerns that to include an evidential presumption in relation to voluntary intoxication would impact negatively on the defendant's right to a fair trial.[124] If a defendant genuinely misjudges a woman's level of intoxication then, subject to s. 1 of the Act and the reasonability of his belief, he will not be convicted of rape, regardless of the existence of the presumption. If, on the other hand, he is unsure of her level of intoxication and suspects that it has affected her capacity to consent, then a presumption

116 Meyer, "Too drunk", n. 106 above, p. 24.

117 Ibid. p. 27.

118 Ibid. p. 27.

119 Ibid. p. 29.

120 Ibid. p. 30

121 Ibid. p. 30.

122 *Bree* appeared willing to accept that this might not be the case but did not take the further step of suggesting that such evidence might instead be corroborative of the victim's lack of capacity to consent.

123 Home Office, *Convicting Rapists*, n. 27 above, at p. 15 and Hansard (HC) 19 November 2002, col. 512.

124 The application and use of the existing evidential presumptions seems to indicate that they are fairly easily rebutted. Of course, this might be an argument against including such a presumption. However, as I argue above, the benefits of such a presumption go beyond the scope of the individual trial.

would be important in reminding the jury that the question to be considered is not only whether the complainant consented but whether the defendant was aware that she had drunk to the level where she had lost her capacity to consent.[125] Finally, as Rumney notes and as argued above, the omission of such a presumption appears to make a distinction between the "deserving" and "undeserving victim"[126] and, thus, feeds into precisely the kind of sexual stereotypes that the reforms to the law were supposed to challenge. As the consultation process showed, the wording of such a presumption would be difficult for the reasons given by the Court of Appeal.[127] However, in the light of the recurrent difficulty with low conviction rates, would it not be better for legislation to err on the side of caution by including the voluntarily intoxicated and thus send out a clear message to society?[128] Given the powerful role of the criminal law in reinforcing appropriate sexual behaviour, this could be an effective educational tool, if combined with an equally effective advertising campaign, which directly sought to tackle victim-blaming attitudes.[129]

The Court of Appeal accepted that leaving consent and capacity to the jury may permit different decisions in similar cases. This is confirmed by Finch and Munro's research, which shows that, in the absence of further definition, mock jurors interpret capacity in different ways leading to different outcomes in similar situations.[130] However, the court justifies this by suggesting that it can be tempered by judicial direction and is in the interests of protecting sexual autonomy and allowing individuals to set their own limits.[131] Of course, whether this is a good thing depends on whose autonomy is being protected and what type of autonomy. The available research on the effects of alcohol on sexual behaviour and rape[132] indicates that a significant number of rape victims were intoxicated at the time of intercourse leading to the conclusion that the use of alcohol is pervasive as a tool in acquaintance rape.[133] Further research indicates that there is a commonplace practice of "loosening women up" with alcohol as a precursor to making sexual advances[134] and that

125 See also Wallerstein, "Drunken consent", n. 91 above, p. 341.

126 Rumney and Fenton, "Intoxicated consent", n. 8 above, p. 288. The deserving victim would not allow herself to get drunk and if she does get drunk then she does not deserve the protection of the law.

127 *R* v *Bree* [2007] 2 All ER 676 at para 35

128 Rumney and Fenton, "Intoxicated consent", n. 8 above, p. 289.

129 Temkin and Krahe, *Sexual Assault*, n. 4 above, pp. 109–23, suggested that the current Home Office campaign had not been effective. However, a more recent campaign by Rape Crisis, Scotland has been evaluated much more positively. The "This is not an Invitation to Rape Me" campaign attempts to confront victim-blaming attitudes in a direct way using a range of images and supporting materials which invite the viewer to examine their attitudes to a range of situations surrounding relationships, drink, dress and levels of intimacy. The stated aim of the campaign is to support women who have been raped, assign responsibility to the (male) attacker and to challenge victim-blaming attitudes. This campaign originally ran in the USA using posters, stickers and public service announcements to attack the perception that when a woman is raped, she asked for, deserved or wanted it. The campaign has also been supplemented in Scotland by a TV advert entitled "Not Ever".

130 Finch and Munro, "Breaking boundaries?" n. 63 above, p. 306; Finch and Munro, "The Sexual Offences Act", n. 63 above.

131 *R* v *Bree* [2007] 2 All ER 676, para. 38. The court refers to the "myriad of circumstances in which the issue of consent may arise". On this point see Gardner, "Appreciating *Olugboja*", n. 87 above, p. 281.

132 HMCPSI, *Report on the Joint Inspectorate into the Investigation and Prosecution of Cases involving Allegations of Rape* (London: HMSO 2002); P Sturman, *Report on Drug Assisted Sexual Assault* (London: Home Office 2000).

133 Finch and Munro, "Intoxicated consent", n. 63 above, at p. 779; N T Harrington and H Leitenburg, "Relationship between alcohol consumption and victim behaviours immediately preceding sexual aggression by an acquaintance" (1994) 9 *Violence and Victims* 315–24.

134 D L Masher and D L Anderson, "Macho personality, sexual aggression and reactions to guided imagery of realistic rape" (1986) 20 *Journal of Research and Personality* 77–89; P Y Martin and R A Hummer, "Fraternities and rape on campus" in P B Bart and E G Moran (eds), *Violence Against Women: The bloody footprints* (California: Sage 1983); HMCPSI, *Report*, n. 132 above; Sturman, *Report*, n. 132 above.

sexual seduction is often linked with drink spiking.[135] Given that the new statutory definition was said to be informed by an understanding that persons engaging in sexual activity have to understand the consequences of their acts and research has shown that such foresight of consequences is unlikely in later stages of intoxication, is this really a matter that could not be successfully dealt with by further legislation possibly coupled with the use of expert evidence regarding the effects of alcohol on a drunken woman's mental capacity to consent?[136] After all, one of the aims of the legislation was to inform the public about the boundaries of acceptable sexual behaviour. This could be supplemented by an effective advertising campaign such as the current Rape Crisis Scotland campaign and the use of expert evidence in relation to alcohol consumption and capacity to further educate juries. As Conaghan and others have noted, the law already recognises some situations when a person's ability to make important choices is significantly impaired and it has been prepared to set limits on what is acceptable in relation to the defence of intoxication for specific intent crimes, so why is this not possible in relation to intoxication and consent?[137]

Further, are the legal principles set out in the Act really so unproblematic, or could they be expressed in such a way as to alleviate some of the problems raised in this type of case? Contrary to the views of the Court of Appeal, academic commentary continues to criticise the legislative provisions as "extremely vague".[138] Following the Court of Appeal decision in *Bree*, the concepts of freedom, choice and capacity continue to be left open to jury discretion and possible prejudice. In the interests of clarity, would it not have been better for the court to endorse a statutory definition of capacity that could build on some of the more positive aspects of the previous case law?[139] Elvin suggests that a possible definition could state that the complainant does not have the capacity to consent where her knowledge and understanding are so limited that she is not in a position to decide whether to agree.[140] However, I would prefer a slightly wider definition which could more clearly encompass those who had been drinking heavily, such as the victims in *Bree*, *Dougal* and *Hagan*, namely victims who were so drunk that they were vomiting, and possibly unable to walk unaided and talk coherently. Admittedly, it is quite difficult to frame such a definition clearly, which means, as Wallerstein notes, that there would still be a wide grey area, but ideally such a definition should take into account that heavy drinking can distort a person's decision-making to the extent that it negates her ability to give meaningful consent.[141] This would make clear to the jury that knowledge and understanding are integral to consent and would take into account the negative as well as the positive aspects of sexual autonomy.

I would argue that returning to the *Olugboja* interpretations of consent allows the law of rape to continue to reflect male perceptions. The judgment in *Bree* clearly reflects the fact that, despite reform, the law has not overcome its difficulties in distinguishing between rape and consensual sex. By detaching consent from its broader context[142] and leaving important issues of definition to the judge and jury, the precise myths and stereotypes that

135 CentreLGS, *Response to the Office for Criminal Justice Reform's Consultation Paper: Convicting rapists and protecting victims of rape – justice for victims of rape* (AHRC Research Centre for Law, Gender and Sexuality, July 2006), www.kent.ac.uk/clgs.

136 HMCPSI, *Report*, n. 132 above; Sturman, *Report*, n. 132 above.

137 CentreLGS, *Response*, n. 135 above.

138 Elliott and de Than, "The case for a rational reconstruction", n. 6 above, p. 238; Temkin and Ashworth, "The Sexual Offences Act", n. 8 above.

139 Scarman LJ in *Lang* (1976) 62 Cr App R 50 stated that "*the critical question is . . . whether the [complainant] understood her situation and was capable of making up her mind*".

140 Elvin, "Intoxication", n. 54 above, p. 155.

141 Wallerstein, "Drunken consent", n. 91 above, p. 332.

142 For fuller arguments on the contextualisation of consent, see Lacey, *Unspeakable Subjects*, n. 39 above.

the Act sought to avoid risk being reintroduced. Further, by situating consent as a matter of proof on the facts of the individual case, the court continues to fail to appreciate the context in which rape takes place. Rather than encouraging a communicative model of consent and shifting the doctrinal boundaries of rape in the direction that the review intended, this return to previous principles could have the effect of undermining the experiences of women who allege rape under such conditions as being "real" rape and fails to take the step of appreciating that sex under such conditions is very different from consensual sex.

The underlying rationale of the judgment would appear to be a reclaiming of judicial discretion in the way in which consent is put before a jury. This allows the judiciary to maintain a strict control of the facts and themes in summing up a case. What this means in practice is that a judge in an individual case can manipulate the discourses of consent to privilege a particular notion of heterosexuality and to reward women who exhibit approved characteristics. In this way, through its strict control of the discourses, the law will be able to continue to police the sexuality of women and maintain the privileged position of male-centred notions of sexual intercourse at the heart of rape law. It could be argued that, despite the research which shows jurors tend to import myths and stereotypes into the jury room and their decisions,[143] in the *Bree* case, they got it right and that their instincts are being undermined by the judiciary. Experience has shown that judges are not adept at providing scope within the discourses of rape trials for a mutuality of sexual desire or women's affective experiences. Through a subtle process of gender construction, the judgment defines and limits female experience to a social and cultural stereotype: namely that women are inclined to lie about allegations of rape and men need protection from women's dangerous sexuality. *Bree* undermines the harm done to rape complainants by unwanted intercourse, subordinates the language of the Act to the common sense of the judiciary, and reinscribes into the law the very myths and stereotypes about heterosexual relations that the legislation was designed to undermine.

Laying the ghosts to rest?

My consideration of the reform of the law on consent would seem to indicate that, whilst the changes in isolation may be a step in the right direction, they have had little effect on their own. This has led to some commentators suggesting that improvement could only be achieved if rape law moves away from a focus on consent.[144] This is a matter that merits further research and consideration. However, I would tentatively suggest that such a move would not necessarily lead to a less one-dimensional understanding of sexual autonomy or recognise the reality that consent cannot be removed from the circumstances in which choice is made. Instead, I would suggest that what is required is, firstly, a more generous definition of consent which allows a much lower level of intoxication to negate the capacity to consent and, secondly, public information campaigns aimed at displacing the gender roles, myths and stereotypes which interfere with the appraisal of specific evidence in cases and undermine the efforts of policy and legislation. To some extent this is already happening and should be encouraged.[145] However, it is a shame that the government

143 Finch and Munro, "Juror stereotypes", n. 63 above; G Chambers and A Millar, *Prosecuting Sexual Assault* (London: Scottish Office/HMSO 1986); Lees, *Carnal Knowledge*, n. 65 above; Kelly et al., *A Gap*, n. 63 above.

144 V Tadros, "Rape without consent" (2006) 26 *Oxford J Legal Stud* 515; M Madden Dempsey and J Herring, "Why sexual penetration requires justification" (2007) 27 *Oxford J Legal Stud* 467.

145 Home Office, *Convicting Rapists*, n. 27 above.

allowed the media and the courts to undermine its recent consultation on rape.[146] The consultation paper, like the National Consent Campaign, was not unproblematic for feminist analyses in that it continued to place the emphasis in rape law on the regulation of male sexuality rather than the protection of women's sexual autonomy. However, as Conaghan and others argued in their response, the symbolic power that an evidential presumption could have wielded should not have been underestimated.[147] One of the stated aims of the SOA was to protect the victim's autonomy. A presumption dealing with self-induced intoxication would have been an important educative tool to indicate to juries (and the public at large) that it is the victim's state and not the means by which it is achieved that is important in determining the point at which consent is negated. This social and cultural role of the law is important as it is not the law in abstract but the law in its social, cultural and philosophical context that has been so effective in undermining rape complainants' experiences. However, as *R* v *Bree* demonstrated, the most well-meaning law reform is always subject to consideration and interpretation by a judiciary still not well versed in issues of gender awareness.[148] Unless and until law reform is considered in such an inclusive manner, effective legal change remains a distant dream.

146 For an example of some of the media comment on the issue, see I Oakeshott, "Men who rape drunk women face tougher law", *Sunday Times*, 17 June 2007; J Clarke, "Men warned to check for consent before sex", *The Independent*, 7 March 2006

147 CentreLGS, *Response*, n. 135 above.

148 For recent research on judicial attitudes to law reform in this area, see Temkin and Krahe, *Sexual Assault*, n. 4 above, chs 6–7.

NILQ 62(1): 119–36

The International Criminal Tribunal for the Former Yugoslavia: paving the way for modern international humanitarian law enforcement

ANDREW WOODCOCK

Cayman Islands Law School

Introduction

The early 1990s was a period which, arguably, saw a revitalisation of internationalism. This was met with varying degrees of success, from the liberation of Kuwait, to the failed attempts to curb the behaviour of Somalian warlords, but importantly, did effect a revival of the peace-keeping system.[1] An important corollary of the perceived need to intervene in the affairs of states in turmoil was the perception of a need to punish those who had caused the turmoil. Although the establishment of a post-conflict criminal tribunal was not entirely new, there were fundamental practical and theoretical difficulties. In modern international law, the ground for the creation of a truly international body was broken with the establishment of the International Criminal Tribunal for the former Yugoslavia (the Tribunal, or ICTY), and it is from a study of the background and early days of the Tribunal that we can see many of the difficulties – legal and political – inherent in the establishment of an international criminal tribunal. Ultimately, however, it can be seen that the historical lessons available to the founders were diligently applied, to produce a body which has evidenced both legal and practical effectiveness.

The period leading up to the formation of the ICTY was one involving significant regional instability – in a region historically renowned for its political volatility. A feature of that instability was that military action took place against both military and civilian targets. Belated public awareness of these actions eventually led to an outcry and demand for some response and, most importantly, justice for those civilians who suffered at the hands of the military. The call for a response resulted in two major steps being taken: military intervention by NATO and United Nations forces, and later, the establishment of a judicial body to judge those individuals who conducted aggressive military action against civilian populations.

The decision to create an ad hoc tribunal to deal with alleged human rights abuses was not without discomfort both for the United Nations, and certain key states which had been involved in the region: primarily France, the United States and United Kingdom. Accusations of inaction, or at least paucity of action, were levelled at all of these entities prior to the formation of the contemplated body.

1 See, for example, "UN peacekeeping record", news.bbc.co.uk/2/hi/health/892592.stm; and M Lacey, "After failures, UN peacekeepers get tough", *New York Times*, 24 May 2005.

This paper will consider the circumstances giving rise to the establishment of the ICTY, and the criticisms which have been raised in respect to the alleged delays and inactivity. It will be contended that, in fact, the criticisms are somewhat unwarranted, as these states were simply endeavouring to deal with genuine practical difficulties confronting them; these were the same difficulties facing their predecessors after both world wars. It will also be submitted that, ultimately, these obstacles were effectively overcome and a credible judicial body was created. The value of this institution lies not simply in its own contribution to justice in the region, but in its status as a template for subsequent international humanitarian courts.

Legal background to the Tribunal

A formal recognition of the legal force of international law was the subject of some substantial discussion amongst theorists for a considerable period of time. The Roman concept of *ius gentium* governed relations between Rome and its conquests and neighbours. Hugo Grotius developed the principles of *ius gentium*, and sought to systematise the law of nations with his *De Jure in Belli ac Pacis* in 1625, which is considered to be the first comprehensive statement of the law of nations. However, even as recently as 1832, John Austin – one of England's most influential legal scholars – refused to accept that international law was anything more than a moral code which ought to act as a guide to the behaviour of states.[2]

Austin's twentieth-century successor and critic, H L A Hart observed that "international law is at present in a stage of transition"[3] towards a level of development which would satisfy the positivists' definition of what constitutes law. When that transition is complete, according to Hart, international law will be a "developed" legal system, and therefore be capable of being considered properly as law.[4] Since Professor Hart propounded that idea, the institutional structure of international law has developed significantly, going well beyond being merely a set of moral guidelines for states.

It is therefore submitted that twentieth-century international law gained a degree of force through its institutional structure (particularly in the form of the United Nations), such that now, in the twenty-first century, even Austin may agree that international law is true "law". The institutions of modern international law now give it a force of compulsion which would be likely to satisfy Austin's desire for a command and sanction. This is rarely seen more clearly in the international arena than in penalising those who breach standards of international humanitarian law.

Antecedents of the Tribunal

It is commonly believed that the Nuremburg and Tokyo trials which followed the Second World War were the first serious attempt to bring to justice those who misconducted themselves in times of war. In fact, this is far from the case and, as early as 1815, there had been moves and discussions in respect to the trial of purported war criminals. The founders of the Tribunal were able to learn much from the lessons offered by these historical examples.[5]

2 J Austin, *The Province of Jurisprudence Determined*, Lecture V (London: Weidenfeld & Nicholson 1954), p. 187.

3 H L A Hart, *The Concept of Law* 2nd edn (Oxford: OUP 1994), p. 236. It is important to note, however, that this observation was first made in 1961.

4 H McCoubrey, "Natural law, religion and the development of international law" in M W Janis and C Evans (eds), *Religion and International Law* (The Hague: Martinus Nijhoff 1999), p. 178.

5 For example, there was an unsuccessful move to prosecute leading Bonapartists after the surrender of France in 1815; Henry Wirz, the officer commanding Fort Sumter, a confederate prisoner of war camp during the United States Civil War, was executed for murder, as a result of mistreatment of prisoners; and several irregular colonial troops were executed by the British while fighting the Boer War.

At the conclusion of the First World War, the Allies required both the Turks and the Germans to try certain of their citizens for crimes committed during the course of the war. In the case of the Turks, the British sought redress against senior Turkish officials for the massacre of large numbers of Armenians throughout the war. Despite best intentions, the national courts-martial put in place to try accused persons collapsed with very few trials completed, and even fewer convictions, as Turkey descended into civil war.[6]

By 1920, Winston Churchill, then War Minister, urged release of those prisoners held pending trial. Even though Churchill was in favour of pressing charges against war criminals – a position which he had maintained throughout the war[7] – he recognised the risk to British troops in holding the prisoners for trial. He therefore conceded that it was better to release the prisoners than risk further British loss of life.[8] This problem of *realpolitik*, in the prosecution of accused persons is an ongoing theme in the field of international humanitarian law, and is certainly one faced by the Tribunal in its early days. The Tribunal relies on the goodwill of individual states for the surrender of accused persons, and at first this was not forthcoming.[9]

The attempted prosecution of German war criminals similarly failed in the period following the First World War. The Allies required the Germans to prosecute some 900 officers and men for various offences which were said to have been committed during the war. Again, the intention was for these offences to have been heard before domestic courts-martial. In fact, there were very few cases heard, fewer convictions, and the lenient prison sentences imposed were substantially truncated. In essence, nothing was achieved through these show trials.[10]

These early experiences highlighted two major difficulties inherently associated with establishing a body purportedly exercising international criminal jurisdiction. These were the same difficulties which were experienced almost 80 years later. The first is the question of independence. On the one hand, it is a major potential weakness of credibility for a victorious nation to establish its own tribunal, because it lacks the outward show of independence, as well as suffering from the appearance of an exercise in vengeance, rather than justice. On the other hand, to leave the prosecution in the hands of the home state is simply to invite lethargy or inaction, as the state is likely to be reticent to prosecute its own.[11]

---

6 G Bass, *Stay the Hand of Vengeance: The politics of war crimes tribunals* (Princeton: Princeton University Press 2002), pp. 134–5.

7 Note, for example, FO383/32/45699, 17 April 1915.

8 FO371/SO90/E10303, 19 July 1920.

9 *Report of the International Tribunal for the Prosecution of Persons Responsible for Serious Violations of International Humanitarian Law Committed in the Territory of the Former Yugoslavia Since 1991*, S/1995/728, 23 August 1995 (hereafter *Second Annual Report*), paras 129, 189, 191.

10 "Commission on the Responsibility of the Authors of the War and on Enforcement of Penalties" (Jan–Apr 1920) 4(1) *American Journal of International Law* 95–154. The commission found that personal liability could be imposed on accused persons, irrespective of rank. However, the commission also found that each belligerent state could try accused persons before their own courts, whether civil or military.

11 Note, for example, the My Lai massacre in Vietnam, in which estimates of up to 500 civilians were killed by United States troops. Notwithstanding his conviction by a court-martial, Lt William Calley, one of the ringleaders of the massacre, only ever spent four-and-a-half months in prison. More recently, the United States passed the American Service-Members' Protection Act 2002, which: permits the US President to authorise any means necessary for the protection of American personnel held by the ICC (s. 2008); and contemplates the withholding of military aid to states which are a party to the court (s. 2007). Notwithstanding that neither of these provisions has ever been acted upon, it demonstrates an unwillingness on the part of the United States to allow its citizens to be the subject of international criminal prosecutions.

As will be seen, this problem has been overcome for the Tribunal in the fact that it was created and managed by the United Nations.[12]

The second major problem experienced is that of enforcement. Churchill's acknowledgment that the capture and retention of potential defendants involved an inherent risk to British troops highlights a problem which continued to be a major issue in the Balkans.[13] In fact, it was arguably even more so, because the conflict itself was entirely unrelated to the states expected to act as police and prison guards. This both heightened the sense of interference for the home state, as well as raising the question of why the "interfering" states were in fact becoming involved, and risking the lives of their troops.

In its *First Annual Report*, the ICTY noted that the Tribunal has been the subject of some criticism because of the lack of compulsive power. Although this certainly was, and remains, a problem, there is an indirect solution available. The mechanism for securing attendance of accused persons is by bringing any non-compliance to the attention of the Security Council, which then has the power to take appropriate steps to assist the Tribunal.[14]

Some of these difficulties were addressed immediately after the Second World War, with the formation of the Nuremburg and Tokyo Tribunals. The ICTY was the first war crime tribunal brought into existence since the end of the Second World War, after the work of the Nuremberg and Tokyo tribunals.[15]

The establishment of the Tribunal was therefore of major importance in the development of international criminal law. There had been a period of over 50 years in which no such tribunal had existed on the international scene, even though there was a much higher level of internationalism than there had been during the years leading up to the Second World War.[16] Furthermore, given the substantial interval of inactivity in this area of law, and the likelihood of the need for future action to be taken in the field, it was of great importance that the international community should "get it right" in the establishment of a tribunal to deal with highly publicised infringements of human rights.

Of particular importance was ensuring that the state parties in the region would sufficiently trust the process of the Tribunal, such that they would refrain from taking matters into their own hands. A stated objective of the Tribunal has always been not only to restore the rule of law to the region, but also restore peace. As the Tribunal itself has commented: "*One of the main aims of the Security Council was to establish a judicial process capable of dissuading the parties to the conflict from perpetrating further crimes.*"[17]

It is submitted that the high profile media attention meant that this was not a passing problem for which a temporary solution could be found.[18]

---

12 *Report of the International Tribunal for the Prosecution of Persons Responsible for Serious Violations of International Humanitarian Law Committed in the Territory of the Former Yugoslavia Since 1991*, S/1994/1007, 29 August 1994 (hereafter *First Annual Report*), para. 10.

13 In fact, it continues to be a problem today, as is demonstrated by the observations of the President of the Tribunal, when he noted that one of the principal functions of the Tribunal is to encourage individual states to exercise jurisdiction over the relevant offences. See P Robinson, "A critical assessment of the impact of the ICTY on the states of the Former Yugoslavia", address to the University of Zagreb, 12–13 June 2009, paras 2–3.

14 *First Annual Report*, para. 17.

15 M McDougal and F Feliciano, *The International Law of War: Transnational coercion and world public order* (New Haven: New Haven Press 1994), p. 704.

16 Ibid. p. 706.

17 *First Annual Report*, para. 11.

18 See R Gutman, "Death camps: survivors tell of captivity, mass slaughter in Bosnia", *Newsday*, 2 August 1992; and E Sciolino, "In Bosnia, peace at any price is getting more expensive", *New York Times*, 10 January 1993.

Historical background to the Tribunal

It ought to be noted from the outset that the Tribunal was not brought into existence to deal with any single conflict. There were, in fact, three major conflicts with which the Tribunal was intended to deal. These were, firstly, the brief conflict between Slovenia and the army of Yugoslavia in 1991, although events subsequently showed that this particular engagement has given rise to the fewest investigations. The second was the war in Croatia from 1991 to 1995. Finally, the war which caused the greatest outrage amongst the international community, and led to the most vocal demand for the creation of a war crimes tribunal, was the war in Bosnia from 1992 to 1995.[19]

In addition, the events in Kosovo in 1998 were later added as the subject of the Tribunal's jurisdiction. However, these were substantially after the creation of the Tribunal, and therefore do not figure in this paper's consideration of the circumstances surrounding its establishment.

Importantly, the nature of each of these conflicts was, in part at least, an ethnic or racially based aggression. The corollary of this very specific underlying cause was that, in addition to action being taken to achieve broad military objectives, the protagonists also engaged in substantial military action against civilian populations. The best known of these objectives was the so-called "ethnic cleansing".[20]

The concept of ethnic cleansing is today so well publicised and well known as to be almost trite. However, this was not the case in the early 1990s when the events were actually occurring and the phrase was being coined. Ethnic cleansing has been described as:

> a purposeful policy designed by one ethnic or religious group to remove by violent and terror-inspiring means the civilian population of another ethnic or religious group for certain geographic areas.[21]

The acts, which were subsequently said to be war crimes, committed in this region were not, however, limited to ethnic cleansing. They involved the widespread killing of non-combatants, rape and the destruction of civilian property.[22]

Although it is today generally accepted that the preponderance of the misconduct was carried out by the Serb forces, military action against civilian populations was one of the primary causes for concern in the region, and that conduct was carried out by all parties.[23] Therefore, the issue of culpability for atrocities committed was not one which could obviously and easily be ascribed exclusively to any one protagonist. This is a point of some interest in respect to the formation of the Tribunal, because it directly impacted upon its credibility, in that it was not a "victors' court" in the sense of the earlier ad hoc tribunals.

In the months following the close of the Second World War, the culpability for causing the war, and subsequent actions during the war, was clear. That was not the case with the Balkans conflicts. There was therefore a high level of credibility to be derived from the establishment of a non-partisan body by an international umbrella organisation, designed to take punitive action against all those who committed international humanitarian crimes, rather than one side of the political conflict. Therefore, the fact the ICTY was not an entity

19 E Sciolino, "US names figures it wants charged with war crimes", *New York Times*, 17 December 1992.

20 A Roberts and R Guelff, *Documents on the Laws of War* (Oxford: OUP 2002), p. 565.

21 *Report of the Commission of Experts Established Pursuant to the United Nations Security Council Resolution 780 (1992)*, S/1994/674, 27 May 1994, para. 130.

22 Article 4(2) of the ICTY statute prohibits this conduct by including it within the scope of the definition of genocide, and Article 2(d) prohibits the destruction of property as one of the "grave breaches of the Geneva Conventions of 1949".

23 S Ricchiardi, "Exposing genocide . . . for what?" (June 1992) *American Journalism Review*.

whose sole *raison d'être* was to punish a losing protagonist added significantly to its value as an independent judicial entity.

Even in the absence of public knowledge of and reaction to these measures, the conflict had already given rise to a determination by the Security Council in September 1991 that the fighting in the region amounted to a threat to international peace and security.[24] This allowed the United Nations to take action under chapter VII of its charter, in the form of the imposition of an embargo on the delivery of weapons and other military goods to Yugoslavia. Most importantly, it also allowed the Security Council to create an organ under Article 29 of the charter to give effect to its resolutions.[25]

The news of the extent of the atrocities occurring in the region first broke in the United States media in July/August 1992.[26] As public awareness grew, there was an increasing level of pressure from human rights groups imposed on the principal Security Council nations, as well as the United Nations itself, to take actions to redress the problems in the Balkans. These events subsequently led to a further resolution by the Security Council on 6 October 1992 for the appointment of a Commission of Experts to investigate the allegations made in the media.[27] That Commission of Experts delivered its report on 10 February 1993. The report confirmed the factual allegations which had already been widely publicised in the world media.[28]

Relying on the recommendations made in this interim report, the Security Council adopted Resolution 808 on 22 February 1993, by which it accepted that a Tribunal ought to be established to investigate, indict and hear charges relating to violations of international law. In order to give effect to this resolution, the Security Council subsequently passed Resolution 827 on 25 May 1993, by which the Tribunal was actually established, and the Statute of the Tribunal was adopted.

Criticisms of the early days of the Tribunal

There has been a lot of criticism of both the time it took to establish the Tribunal, and subsequently for the Tribunal itself to take any action to give effect to its mandate. It has been suggested that this was not an important political issue for the Bush administration at the time, and any action which was going to occur was only likely in response to public pressure.[29] This is, of course, what happened when it became apparent that there were events such as ethnic cleansing and genocide taking place in an area for which the United States was already partially responsible, through its participation in NATO (North Atlantic Treaty Organisation). It was therefore necessary for groups interested in human rights to exert pressure to make the issue sufficiently political for action to be taken in the Balkans.[30]

The Balkans was not the major consideration for the Bush administration. By the time these events came to the public eye, the First Gulf War had been over for only a short time,

24 Resolution 713, 25 September 1991. The resolution also called on the parties in the region to refrain from their continued breaches of a cease-fire agreement.

25 Article 29 provides that: "The Security Council may establish such subsidiary organs as it deems necessary for the performance of its function."

26 The story was originally brought to light by *Newsday* journalist Roy Gutman in a series of articles from 19 July 1992: "Prisoners of Serbia's war", *Newsday*, 19 July 1992; "Like Auschwitz", *Newsday*, 21 July 1992; and "Bosnia's camps of death", *Newsday*, 2 August 1992.

27 Resolution 780 (1992).

28 See, in particular, *Report of the Secretary-General on the Activities of the International Conference on the Former Yugoslavia* (S/25221), 2 February 1993, Annex I, paras 1–4.

29 B Magas and I Zanic (eds), *The War in Croatia and Bosnia-Herzegovina* 1991–1995 (London: Frank Cass), p. 273.

30 *Interim Report of the Commission of Experts Established Pursuant to the United Nations Security Council Resolution 780 (1992)* S/25274, 10 February 1993, paras 5–7.

and threats were being made by the United States and British governments to bring Saddam Hussein before a war crimes tribunal, albeit that such threats were relatively abstract. At a press conference on 1 March 1991, President Bush refused to rule out the possibility of a prosecution of Saddam Hussein.[31]

The benefit for the United States and United Kingdom in bringing that tribunal into existence would have been twofold: firstly, the war was over and there would have been no political complications in creating such a body; and, secondly, given that the war was over and won, there would have been minimal risk to British and American troops in giving effect to the determinations of such a tribunal. There was substantial discussion in the media of the possibility of pursuing such a course. Ultimately, it even became a significant electoral issue in the 1992 US presidential election. However, whether it was because of simple lethargy, or a desire to be free of any further involvement in the territory, or even that it was no longer politically expedient to do so, neither Hussein nor any of his lieutenants were brought before any court to answer for their actions until many years later.

Bass, in his very useful account of war crime trials throughout history, argues that it was a lack of political will on the part of the new US President, Bill Clinton, which resulted in a degree of sluggishness in response to the problems in the Balkans. In particular, Bass cites the President's concerns about committing troops to the region.[32] However, this critique of the alleged *realpolitik* of the new American leader is contradicted by the significant material contribution made to the establishment of the Tribunal by the United States. In addition to making the largest single voluntary financial contribution to the Tribunal, the United States also provided 22 professional staff on secondment.[33]

It is submitted that, in promoting the Tribunal, the United States administration must have been aware that it was creating an institution which would require "police" services. If the political will were entirely lacking, as Bass suggests, the United States would not have made a material contribution which effectively ensured the survival of the Tribunal. It is certainly accepted that there was a potential political backlash for the new President in committing troops to a foreign conflict. However, the Democrats had made so much of the failure by the Bush administration to prosecute Saddam Hussein that for Clinton to fail to act in the first months of his presidency would have been just as damaging politically.[34]

Both Britain and France have been subject to particular criticism. This was both in the lead-up to the establishment of the Tribunal, as well as in its immediate aftermath. Britain has been accused of providing virtually no assistance in the creation of the Tribunal; giving no money towards its initial funding; and providing only one staff member. This is in contrast to the United States, whose initial contribution was of substantial funds and the provision of some 20 of the Tribunal's initial 60 staff. This latter point did, however, lead to a criticism that the Tribunal was dominated by an "American mafia". This would appear, however, to be a somewhat unfair criticism, given the fact that no other members of the United Nations seemed willing to be quite so forthcoming in support of the Tribunal.

The French were subject to substantially the same criticism as that levelled against the UK. Both were accused of "foot dragging" in the lead up to the formation of the

31 "After the war: the White House; excerpts from Bush's news conference on post-war plans", *New York Times*, 2 March 1991.

32 Bass, *Stay the Hand*, n. 6 above, pp. 214–15.

33 *First Annual Report*, para. 186.

34 See a series of articles by the *New York Times*' Elaine Sciolino: "The 1992 campaign; Bush's greatest glory fades as questions on Iraq persist", 27 June 1992; "US is said to withhold evidence of war crimes committed by Iraq", 6 July 1992; and "Iraq policy still bedevils Bush as congress asks: were crimes committed?", 8 August 1992.

Tribunal. As with the British, France contributed very little funding, and failed to contribute any staff at all.[35]

One important explanation for the delay by the British and French governments is their historical experience of the Balkans. The region has long been considered to be a problem for European troops. Historically, it has proven to be one of highly complicated political relationships, and it has been very easy for troops to become embroiled, at great cost to a nation in financial and human terms, without any apparent result, or ease of withdrawal.[36] In these terms, resisting involvement in foreign and unrelated conflict is, to a large extent, understandable. The concerns expressed by Churchill in 1920 were no less relevant and applicable in 1993.[37]

It would appear to be the case that until the full extent of the problem began to be known in the West, the mere existence of a local conflict was not sufficiently shocking for substantial pressure to be brought to bear on governments.[38] In this respect, contrast the end of the Second World War, at which time the full extent of the Nazi activities became known and there was an immediate call for action. In those circumstances, there was both a personal aspect, in that Germany was in fact the enemy, as well as a growing awareness of the scale of the misconduct. At the early stages of the Balkan conflicts, there was simply not this level of public awareness and, therefore, the task was left to the media and human rights groups to bring the matter to the attention of states, and to the public, who would then apply the necessary political pressure to governments.

However, this response to the issue seems, with respect, a little simplistic and self-serving. The issue of national self-interest ought not be forgotten. Humankind is inherently self-interested, even in its international relations, although that self-interest must necessarily be tempered by the need to live harmoniously as a community of nations.[39] In the period following the Second World War, there was an obvious motivation for prosecuting the vanquished: the desire to wreak vengeance on those who had plunged the world into conflict for the second time in 30 years.

No such personal motivation was at work on the United States, Britain or France in 1993. In fact, the negative political consequence was a factor actively working against those states contributing to establishing a war crimes tribunal. It is the function of the law, and the law-makers, to encourage people to overcome their naturally selfish tendencies.[40] Thus, the delay on the part of these states is entirely understandable. Perhaps most importantly, when reminded of their duties within the international sphere, they ultimately acted and fulfilled their responsibilities as members of the international community.

However, there is a counterpoint to this proposition. Even in the days prior to the establishment of the Tribunal, the United States sought to effect a peace settlement which would have potentially obviated the need to commit troops.[41] Given the natural human tendency towards self-preservation, it was considered highly unlikely that a peaceful resolution could be reached if there were a serious possibility of legal culpability for war crimes being directed at one or more of the parties involved. Furthermore, Slobodan

35 *First Annual Report*, paras. 184–6. Notably, this carried on into the second year of the Tribunal, although Britain and the United States continued to make material contribution.

36 G Campbell, *The Road to Kosovo: A Balkan diary* (Boulder: Westview Press 1999), p. 116.

37 See n. 8 above.

38 Bass, *Stay the Hand*, n. 6 above, p. 211.

39 This is the basis on which Hugo Grotius proceeded in *De iure Belli ac Pacis*, Prol., paras 8 and 10.

40 Aristotle, *Nichomachean Ethics*, Bk 7.

41 J F Burns, "Balkan war trials in serious doubt", *New York Times*, 26 April 1993.

Milosevic, in a letter from the Prime Minister of Yugoslavia to the Secretary General of the United Nations, made it clear that the establishment of any tribunal was yet another example of the victimisation of Serbia.[42] It has therefore been suggested that only by keeping a morally neutral tone to matters relating to this conflict could this peaceful resolution be successfully brought about.[43]

The corollary of that suggestion is that the morally neutral aspect was easier to achieve for the British and French, whose troops were present in the region at the time, and were less partisan in favour of any faction (particularly the Bosnians). Britain and France were more likely to see all parties involved in the dispute as being equally culpable, and therefore, the necessity to institute a tribunal was not as pressing at that time as there was no obvious need to redress a significant injustice.

In addition to this fear was the simple practical fact that no judicial body could have been established in this region without troops to give it effect, both in terms of arresting potential defendants and enforcing its determinations. It was acknowledged that simply establishing a judicial entity would not be sufficient. A military element was a necessary accompaniment to the establishment of the Tribunal. In order to arrest, try and perhaps even give effect to the decisions of the Tribunal, the use of force against potential defendants was always a distinct possibility. In the absence of any United Nations police force, responsibility for this military action would lie with the nations whose troops were present in the region. This possible need to commit further troops gave rise to a degree of French and British resistance to the idea of setting up a Tribunal, with the knowledge that its establishment could result in the deaths of their troops. There were, therefore, immediate and understandable reasons – based on hard historical experience – to explain the British and French resistance to becoming involved in the creation of a war crimes tribunal.

There were similar practical, politically based, reasons for the reluctance of the United States to act. In the context of the domestic political landscape of the United States, these events were occurring within the early months of the Clinton presidency. It was obviously important to any administration not to place United States troops in a position of potential danger. The resistance to action at this early stage was therefore explicable from the perspective of US internal politics.

On the same line of reasoning, yet perhaps adopting a more cynical approach, is the criticism that the major powers – the United States, Britain and France – had a desire to maintain the status quo.[44] For the United States, this was a means of avoiding any risk to its troops by committing them to an unstable region. For France and the UK, the preservation of the status quo essentially meant preserving their respective positions within the European balance of power, by neither weakening their own military power, nor by propping up any new military force on the European stage.[45] Balanced against this political desire were the legal obligations associated with membership of the European Community. In a statement made on 1 February 1993, the European Community confirmed its support for the peace process in the Balkans. In particular, in reference to the United Nations' measures to secure peace, it was said: "The European Community and its Member States are prepared to contribute actively to this end."[46]

42 UN Doc. S/25801, 19 May 1993.

43 Magas and Zanic, *The War in Croatia and Bosnia-Herzegovina,* n. 29 above, p. 274.

44 Ibid.

45 Ibid. p. 275.

46 *Report of the Secretary-General*, n. 28 above, para. 20.

This argument would have it that these powers wanted to contain the conflict and let it burn itself out.[47] They were prepared to maintain a watch on the conflict, ensuring that it did not spill beyond its original borders, but not themselves become sufficiently involved to bring the conflict to an end. In purely domestic political terms, such an approach is highly beneficial to the major powers – and probably more importantly, to the governments of the day, by ensuring that no lives are lost as a result of involvement in another nation's civil war. Of course, the failing of this policy is that it does not take into account the human cost. This safe course would appear to be the one which was initially taken. However, subsequent events overtook the major powers, and their own internal political forces, driven by public opinion, shifted, and forced them to act. In particular, the Tribunal has noted the fact that, in its first year, there were over 600 media articles published throughout the world, and there was a strong element of disapproval for not having achieved more.[48]

Although Britain, France and the United States were the major international players in the region in the early 1990s, there were also other states involved at the time (at the very least at the level of the Security Council), and they too appeared to be reluctant to see the establishment of the Tribunal; if not in express terms, at least in terms of a failure to act. These were nations which broadly come into the category of those with a dubious history of human rights of their own, including, for example, Russia[49] and China.[50] Naturally, such states were resistant to the idea of a precedent being set through the establishment of a "modern" tribunal addressing breaches of international human rights law.[51]

The Tribunal was created at a time when an International Criminal Court (ICC) was also contemplated, but before it had come into existence. There is scope for a strong argument that both China and Russia, in light of their less than unblemished human rights records,[52] were, naturally, both reluctant to assist in establishing precedent in the international community for the prosecution of breaches of humanitarian law. However, any attempt at active resistance would have been transparent and almost certainly would have resulted in negative international press, in the same way that the United States' resistance to the ICC garnered negative response.[53] Therefore, in a similar response to that of France and Britain, there was simply a failure on the part of these states to take any active steps in support, and, given that the backing of the leading nations involved in the region was essential to the formation of the Tribunal, this failure effectively acted as a de facto discouragement.

Similarly, the United Nations itself was the subject of criticism for its failure to take positive action immediately upon the circumstances surrounding the conflict becoming known in the public forum. These criticisms included an allegation of substantial under-funding, in view of the purpose the Tribunal was intended to serve. As an illustration of this point, the initial half-year budget of the Tribunal was $5.4 million, whereas the 2009 budget was $290 million.[54] The limitations and impermanence of initial funding caused the Tribunal substantial difficulty and inconvenience in its early days, in the most basic areas of

47 Bass, *Stay the Hand*, n. 6 above, p. 212.

48 *Second Annual Report*, para. 164ff.

49 Noting Russia's involvement in Chechnya and bearing in mind its own relatively recent invasion of Georgia.

50 Also noting China's involvement in Tibet.

51 It is also interesting to note that, to this day, Russia, China and the United States have still not become parties to the Treaty of Rome, which institutes the ICC.

52 Note Human Rights Watch, *Violations of Humanitarian Law and Human Rights in the Georgia–South Ossetia Conflict*, 1 April 1992; Human Rights Watch, *Human Rights Watch World Report 1998*, 3 December 1997.

53 Human Rights Watch, *United States Efforts to Undermine the International Criminal Court; Article 98 Agreements*, 2 August 2002.

54 *Report of the International Tribunal for the Prosecution of Persons Responsible for Serious Violations of International Humanitarian Law Committed in the Territory of the Former Yugoslavia Since 1991*, S/2010/413, 30 July 2010, para. 94.

obtaining premises and appointing a chief prosecutor.[55] Arguably, this clear disparity between the practical necessity and the initial budget is attributable not so much to a deliberate attempt to detract from the effectiveness of the Tribunal as a misunderstanding of the depth of the problem, and the extent of the needs of the Tribunal. This is supported by the proposition that once the decision was taken to establish the Tribunal, the parties involved were committed to its success. The success or failure of the Tribunal directly reflected upon them. Therefore, it could be said that the failure to ensure adequate resources in terms of staffing was a reflection of ignorance of the full scope of the problem, rather than any more sinister attempt to defeat the purpose of the Tribunal. Ultimately, this argument is borne out by the subsequent substantial increase in funding and resources.

Structure of the statute of the Tribunal

The statute of the Tribunal was adopted by resolution of the Security Council on 22 May 1993.[56] The statute establishes the jurisdiction of the Tribunal, both in temporal and territorial terms. Article 1 of the statute provides that the Tribunal has jurisdiction to hear matters arising out of the territory of the former Yugoslavia from 1991 onwards. Importantly, this article refers generally to the prosecution of any person who has committed a relevant offence. In contrast, the Charter of the International Military Tribunal at Nuremberg referred specifically to the prosecution of the "major war criminals of the European Axis".[57]

Article 2 defines the legal jurisdiction of the Tribunal, which is to hear matters in relation to grave breaches of the Geneva Conventions of 1949. These are described with some specificity within the statute (although, notably, with less detail than is to be found in Article 8 of the Rome Treaty), and include such matters as unlawful deportation. Similarly, Articles 3 to 5 deal with violations of the customs of war, genocide and crimes against humanity respectively. Each area of criminality is defined and described in some detail. Again, this is in contrast with the Charter of the Nuremberg tribunal, which describes the relevant offences in general, inclusive terms.[58]

In contrast with other international judicial bodies, such as the International Court of Justice (ICJ), the Tribunal, through Article 6, has power to deal with natural persons, as opposed to being limited to dealing with states, or entities with an international personality.[59] Article 7 goes on to impose liability on natural persons who not only commit the offences referred to within the statute, but also on those who plan, instigate, order or abet the offences. This latter provision is in order to give effect to an earlier Security Council resolution to the effect that individuals who commit offences of this nature are to be held personally liable for their conduct.[60]

It is interesting to note at this point that there is no need under either the Security Council resolution, or the statute itself, that there be any submission to the jurisdiction of the Tribunal. This is in sharp contrast to both the ICC and the ICJ, both of which require the member state to accede to the jurisdiction of the court before it is entitled to hear matters before it.[61] To this extent, the Tribunal's compulsory, albeit limited – jurisdiction

55 *First Annual Report*, paras 28–51.

56 Resolution 808.

57 Charter of the International Military Tribunal at Nuremberg, Article 6.

58 Article 6.

59 Such as the International Committee of the Red Cross.

60 Resolution 764.

61 Note Articles 12 and 36 of the statutes of the respective courts, although submission to the ICJ is effectively deemed as a consequence of membership of the United Nations.

is much closer to that of the Tokyo and Nuremberg tribunals, for which there was no necessity to submit to the jurisdiction. This is also the reason that the armed forces of independent nations were required to assist the Tribunal, and bring accused persons before it: the home states of the accused persons did not necessarily consent to their citizens being charged.

Once again, this is an illustration of the manner in which the founders of the Tribunal benefited from the historical lessons available to them, and produced an entity which amounted to a pragmatically effective institution, while still an appropriate legal compromise. Given the nature of the accusations being made by all parties involved in the various conflicts, it is almost certain that no state involved would have been willing to consent to the jurisdiction of the Tribunal. Thus, the Tribunal would have been ineffective, or at least substantially less effective. Therefore, the imposition of a compulsory jurisdiction by an independent entity (the United Nations) has the effect of ensuring that a practical result is achieved, while justice is also served.

Article 9 recognises a concurrent jurisdiction between the Tribunal and domestic courts. However, it goes on to provide in Article 9(2) that, where there is such concurrency, the Tribunal has primacy. Therefore, from a procedural standpoint, an accused person ought not be brought before two bodies for the same offence. However, the potential weakness of this provision is that it relies upon the goodwill of the national courts for its efficacy. Similarly, Article 10 states that a person shall not be charged before a domestic court for a matter already dealt with before the Tribunal – effectively incorporating a double jeopardy protection for potential defendants.

These latter provisions are clearly essential for the effective administration of justice, and, perhaps more importantly, the transparent appearance of justice. The argument has been made that a potential danger inherent in the concept of universal jurisdiction is the possible political motivation of prospective prosecutors in determining whether or not to bring matters before local courts.[62] For this reason, granting primacy to a court exercising an international jurisdiction, established by an independent arbiter, and constituted by an international bench protects the rights and interests of the individuals who may be brought before it.

Contrast with the Nuremberg and Tokyo tribunals

While considering the statute of the Tribunal, it is perhaps of interest to note the extent to which the Tribunal was different from the earlier ad hoc tribunals established in Nuremberg and Tokyo after the Second World War. Clearly, there were lessons learned from the establishment and conduct of those bodies which, if not directly applied, were at least understood when the Tribunal was being founded.[63] For that reason, it is interesting to examine the areas in which there is divergence, and note what may have been considered to be the failure of the earlier entities, and the improvements in the later one.

Firstly, and perhaps most importantly, is the fact that the Tribunal is a body of a truly international nature. The 11 judges and the prosecutor of the Tribunal were selected from member states of the United Nations, which gave a potential body of almost 200 nations from which these officials could be selected. This is in direct contrast with the four nations – Russia, France, United Kingdom and the United States – from which the prosecutors and

62 Throughout the *First Annual Report*, substantial commentary is made on the divergences between the ICTY and its predecessors. However, clearly, those early lessons informed the formation and structure of the Tribunal.

63 Roberts and Guelff, *Documents*, n. 20 above, p. 567.

judges were selected for the Tokyo and Nuremberg tribunals.[64] The post-Second World War tribunals can, therefore, be subject to the criticism of being little more than vengeance tribunals, in which the victors sought to exact retribution, rather than justice.

This latter proposition is one which goes to the question of the total effectiveness of the body, which is a feature which cannot be understated. For any judicial entity, its effectiveness is, to a very large extent, reliant upon the perception held by the public, and most importantly by those appearing before it. The fact that the Nuremberg and Tokyo tribunals were established by the four leading victorious nations of the war was a feature which detracted substantially from their credibility, because it reflected the limited perspective of the small number of legal systems from which the key players could be drawn. Furthermore, the Charter of the International Military Tribunal at Nuremberg expressly excluded challenge to the membership of the panel.[65]

This leads to the second divergent feature of the Tribunal, which was that it was established by an independent umpire during the conflict, and therefore covered both actual and prospective criminal offences.[66] The Tribunal therefore reflected an attempt to constitute both a body hearing trials of matters which had already occurred, as well as to amount to a deterrent for future misconduct. The extent to which this was successful will be considered further below.

Under the statute of the Tribunal, rape is expressly included as a crime against humanity.[67] Similarly, genocide is also expressly included as an offence capable of being tried by the Tribunal.[68] These are offences which were included within the Geneva Conventions of 1949.[69] Those conventions were adopted with the benefit of hindsight after the atrocities identified during and after the Second World War. The experience from the earlier bodies was of assistance to the founders of the Tribunal, in that there was a corpus of established crimes, which were recognised as being offensive to international law and were engrossed into the statute as part of its jurisdiction.

Nevertheless, the Tribunal's scope was slightly less expansive that that of the post-war entities. The post-war ad hoc tribunals tried the relatively nebulous offence of "crimes against peace".[70] The rationale at the time was that all resort to war is unlawful. Therefore, Germany, as the aggressor in the war, committed a crime against the preservation of peace. The jurisdiction of the Tribunal is exclusively defined by the statute, and there is no reference to any offence of the nature of "crimes against peace".[71] This could be said to be a return to the post-World War One position which rejected the prospect of criminal liability for acts of aggression.[72]

There are two possible explanations for this limitation on the jurisdiction of the Tribunal. The first is that, although the Serbs were arguably the more aggressive party

64 See Article 1 of the Charter of the International Military Tribunal at Nuremberg.

65 Article 3.

66 As opposed to the Nuremberg tribunal being established by the Agreement for the Prosecution and Punishment of the Major War Criminals of the European Axis (London Agreement), 8 August 1945.

67 Article 4.

68 Article 5.

69 McDougal and Feliciano, *International Law of War,* n. 15 above, p. 706; however, Article 2 also deals with "grave breaches" of the Geneva Conventions of 1949.

70 R Jackson, *Report of Robert H Jackson, United States Representative to the International Conference on Military Trials: London, 1945,* 25 July 1945 (Washington: USGPO 1949), pp. 381–4.

71 The scope of the Tribunal's jurisdiction is exhaustively established in Articles 2 to 5. There is no mention in any of those articles to any reference approaching "crimes against peace", or "crimes of aggression".

72 "Commission on the Responsibility of the Authors of the War", n. 10 above, p. 119.

involved in the various conflicts, there was no clear and obvious aggressor initiating the disputes. This is in contrast with the Second World War, in which it was universally accepted that Germany and Japan were the initiators of the conflict. Although history does tend to bear out this point, it was not even an issue at the time. It was simply presumed.

The second possible explanation can be based around the individual responsibility aspect of the Tribunal's function. The primary purpose of the Tribunal was to bring to justice individuals who had committed specific offences, and hold them responsible for their actions. The concept of a crime against peace is one which is viewed on a higher scale, in that it is directed more at the state's conduct, rather than at that of the individual. For that reason, the offences which are specifically included in the statute of the Tribunal more adequately satisfy the Tribunal's function of bringing individuals to justice for their misconduct. Other tribunals exist to bring states to justice for their misdeeds.[73]

Procedure

Unlike the Tokyo and Nuremberg tribunals, the ICTY had a clearly defined procedure from the outset, embodied in the statute itself. Importantly, this procedure was arrived at after an extensive period of consultation with jurists from a wide variety of jurisdictions.[74] It was intended that the procedure adopted by the Tribunal should take account as much as possible of domestic principles in both the continental European system as well as the Westminster tradition. This is reflected in Article 15 of the statute in which the rules of procedure and evidence are to be found.

In divergence from the rules of the Tokyo and Nuremberg tribunals, no death penalty is available to the judges of the Tribunal.[75] This is reflective of the United Nations' general move away from the imposition of the death penalty and is consistent with such international instruments as the Universal Declaration of Human Rights and the International Convention on Civil and Political Rights.[76] This does, however, raise two important and interesting practical questions for consideration.

The first is in relation to accommodation of convicted prisoners. If the maximum sentences which could be imposed under the statute are terms of imprisonment, then there must be appropriate accommodation for the prisoners serving those sentences. To date, this has been provided by the Netherlands, which has devoted a portion of one of its prisons to United Nations prisoners. The second point is that, under Yugoslav law, the death penalty is available for many of the offences which are within the Tribunal's purview. However, it is again reflective of the primacy of the Tribunal's jurisdiction that this is nevertheless not adopted by the Tribunal in its sentencing protocols.

Finally, under the Tribunal, there is an appeals process available to hear both interlocutory procedural matters and substantive appeals from both conviction and sentence.[77] This is once again a substantial divergence from the operations of the earlier

73 Such as the ICJ and the International Human Rights Commission, although neither of those bodies exercise a criminal jurisdiction per se.

74 In contrast, a very simple procedure is set out in Article 24 of the Nuremberg Charter. The simplicity of this procedure would appear to reflect the inexperience of its draftspeople with the production of rules of international tribunals.

75 Article 27 of the Charter of the International Military Tribunal permits the imposition of the death penalty.

76 As an aside, it is also interesting to note that this is what would have prevented Iraq, or the United States, from executing Saddam Hussein, and it was therefore necessary for a court exercising domestic jurisdiction to pass sentence of death against him.

77 Article 25 of the statute of the Tribunal.

tribunals, whose determinations were final, with no avenue of appeal available.[78] Again, this adds to the credibility of the Tribunal as a body exercising judicial power, rather than simply an instrument of vengeance.

Limitations of the Tribunal

Notwithstanding these theoretical benefits available to the Tribunal from the source of its legal power, there have been a number of practical limitations, some of which have been canvassed already. These limitations have detracted substantially from the capacity of the Tribunal to operate effectively as a body exercising judicial power.

Perhaps the most important is in respect to bringing accused persons before the Tribunal. This was one of the more difficult features of the Tribunal's operations in its early days. In 1996 and 1997, there was substantial criticism levelled against NATO and United Nations troops for a failure to bring persons under indictment before the Tribunal. This was often the case when it was, in fact, known where those persons were. One commentator has noted that various indictees were known to NATO forces, and sometimes even socialised at the same venues as those forces, but were not arrested.[79]

It has been suggested, from the perspective of the United States, that there were two major reasons for this failure to take any real steps towards the arrest of accused persons. The first is the recent memory of the Somalia debacle, in which a number of United States military personnel were killed or injured in an attempt to bring local warlords to justice, ultimately achieving little benefit for the local populace. The risk of involvement of troops in a similar situation was acknowledged, and obviously there was a hope to avoid loss of life with the same apparent lack of result. In the same vein was the fact that it was recognised from public surveys that as many as 70 per cent of the United States population were opposed to the commitment of their troops in Bosnia.[80] An increase in the commitment of troops to the region is almost certainly what would have been required to effect the arrest of a large number of those persons under indictment from the Tribunal.

The other major failing of the Tribunal in its early stages was that generally it was only low and middle-level personnel who were indicted and brought before it. At least in the early days, there was no attempt made to indict senior officers or politicians from either side of the engagements. (This has, of course, subsequently been changed, with the most notable being the indictment and trial of Milosevic: that was considered to be a major step forward in the life of the Tribunal.)

It has been suggested that this failing was a symptom of the fact that the Tribunal was attempting to deal with matters in the course of an ongoing conflict. The argument which has been used to justify the failure to take action against senior officials is that there were attempts underway to broker peace with those senior officials.[81] It would have substantially detracted from the capacity of the United Nations to achieve this end of a brokered peace if it were known that, as soon as the peace were achieved, those involved at the highest level would be indicted for serious offences, involving potentially lengthy prison sentences.

There is some validity to this argument, in theory. It is difficult to envision a circumstance in which a senior politician would be prepared to bring an end to a conflict if it were known that such an act would expose him or her to a charge of crimes against

78 Article 26 of the Charter of the International Military Tribunal provides that: "The judgment of the Tribunal . . . shall be final and not subject to review."

79 Campbell, *Road to Kosovo*, n. 36 above, pp. 113–14.

80 Ibid. p. 116.

81 Ibid.

humanity. However, whether that theoretical argument can be transposed into reality, given the relatively limited number of senior personnel who have been charged even to this day is somewhat questionable. Conversely, it may be said that, by charging a large number of low and mid-level personnel, and securing convictions, the Tribunal is at least ostensibly justifying its existence.

Nevertheless, this does raise an interesting issue, which should be considered. That is, whether there is any basis for the suggestion that a tribunal of this nature should only be formed at the conclusion of a conflict, rather than during the course of that conflict. On the side of favouring such a tribunal at the earliest point is that it allows the tribunal to avoid accusations of being created solely for the purpose of retributive justice. This is, of course, the accusation which was levelled regularly against the Nuremberg and Tokyo institutions.[82] It was one from which they suffered substantially and – in light of other evidence, such as it being a "victor's court" – is quite a reasonable accusation. As a result, those tribunals experienced loss of prestige and acceptability as bodies exercising appropriate jurisdiction. Therefore, there is a case for the tribunal to be established at the earliest time.

However, the contrary argument also carries significant weight. That is that the creation of an entity of this nature during the course of the conflict does little more than complicate the political situation. That is borne out in the present case, in the argument referred to above that the capacity to bring about a negotiated settlement was complicated by the spectre of criminal charges potentially looming against those with whom the United Nations was negotiating. Similarly, there would appear to be little or no evidence to suggest that the formation of the Tribunal did anything to curtail the very conduct which it was established to investigate. This would therefore suggest that entities of this nature are left with little more than retributive justice, if they cannot be said to fulfil any prophylactic purpose.

The Tribunal versus the ICC

It is to be noted that it is only relatively recently that the Statute of Rome, by which the ICC is brought into existence, has come into effect. That is there have now been the requisite number of ratifications for the statute to come into force and for the court to be created. Some thought must therefore be given to the issue of whether, and if so to what extent, this court will oust ad hoc entities such as the Tribunal.

Part 2 of the Statute of Rome sets out the jurisdiction of the ICC. Most relevantly, Article 5(1) states that the "jurisdiction of the Court shall be limited to the most serious crimes of concern to the international community as a whole". This is important to the present context, in that it then goes on to include genocide, war crimes, crimes against humanity and crimes of aggression in that category of serious crime. These terms are more clearly defined and delineated throughout the rest of Part 2 of the statute. The offences are defined in much more detail than in the statute of the Tribunal. However, the point of Article 5 is that, on the face of it, the ICC will not entirely obviate the need for ad hoc bodies such as the Tribunal.

The Tribunal's jurisdiction is not limited to "serious" offences "of concern to the international community as a whole". Its purpose is to try breaches of the Geneva Conventions of 1949, albeit that it limits that jurisdiction to "grave" breaches of those conventions. Nevertheless, a consideration of the scope of the jurisdiction of the Tribunal shows that the level of seriousness of the offences with which it is concerned is not as

82 H Fox and R Meyer, *Effecting Compliance* (London: British Institute of International and Comparative Law 1993), p. 3.

great as those which the ICC handles.[83] Although the latter body is intended to deal with individuals, it is nevertheless exercising a jurisdiction more on the international scale of misconduct than that of the Tribunal. Therefore, it can be said that, notwithstanding the creation of the ICC, bodies such as the Tribunal will nevertheless not be likely to disappear in future.[84]

Conclusion: a promise fulfilled?

It can be seen that the Tribunal was a body created in circumstances of some significant controversy. The historical conditions in which it was formed were characterised by substantial political volatility, in a region in which political instability was endemic. In these conditions, public pressure was placed upon key players in the region: the United Nations, the United States, France and the United Kingdom to take action to restore stability and exact justice from those who had committed serious crimes. Hence, in addition to pressure being imposed for the undertaking of military intervention, pressure was also exerted for the establishment of a judicial body to try such offences. This was the first such ad hoc tribunal since the close of the Second World War.

One of the stated objectives of the Tribunal from the very outset was assisting in the preservation of peace within the region.[85] For all of the reasons already discussed, it would be unforgivably naive to believe that a judicial institution could achieve this end unaided. The participation of NATO and United Nations troops in the region was essential to the restoration of stability. Nevertheless, the Tribunal has gone some significant way towards assisting in the preservation of that stability, and it is contended that this is for two principal reasons.

The first is found in the imposition of penalties on those who have committed offences in the region. It is submitted that this has secured the perception of justice for those who suffered at the hands of the perpetrators. This must necessarily have had the effect of leading the victims away from taking any "self-help" action. The fact that there was an independent entity securing justice for the victims can be said to have taken away at least some of the sting of the acts themselves.

Secondly, and perhaps more importantly, the Tribunal has worked with the nations in the region to ensure that justice is served at both a domestic and international level. In particular, the Tribunal has gradually been transferring skills,[86] and ultimately investigative files, to the prosecutors of the states within the region.[87] This restabilisation of the judicial process in the Balkans has the capacity to aid in achieving both political and military security, by ensuring that each individual state takes a degree of responsibility for maintaining peace and good order. Although this may not be a perfect solution to the problems of the area, it is certainly one which has been shown to be effective thus far, and also represents a somewhat unique use of a judicial institution to achieve international political order.

It is fair to say that in responding to the pressure, those parties satisfied many of the theoretical requirements for such a body. The lethargy with which the Tribunal was

83 Fox and Meyer, *Effecting Compliance*, n. 82 above, p. 7.

84 Note that this is borne out by the creation of the tribunal for Rwanda.

85 *First Annual Report*, paras 7 and 11.

86 *Report of the International Tribunal for the Prosecution of Persons Responsible for Serious Violations of International Humanitarian Law Committed in the Territory of the Former Yugoslavia Since 1991*, S/2009/394, 31 July 2009, para. 76ff.

87 *Report of the International Tribunal*, n. 54 above, para. 71ff and 80–5.

established, the paucity of funding at the outset and the absence of any serious attempt in the early years to bring accused persons before the Tribunal have all been said to be indicative of a lack of commitment by the parties involved. Initially, it could be said that they were found wanting in many practical aspects, although this seems to be more a symptom of both seeking to protect their own citizens, as well as misapprehending the enormity of the task, rather than deliberately avoiding the leadership responsibility which they took upon themselves. Ultimately, the theoretical lessons from the Nuremberg and Tokyo tribunals had been well learned, as well as the practical lessons from even earlier attempts to try soldiers for criminal conduct during time of war. The statute of the Tribunal was crafted in such a way that the status and prestige of the Tribunal as an independent judicial entity were preserved, while still imposing personal liability on those individuals who committed atrocious offences. In the end, it is clear that compromise was essential, in order to achieve any effective outcome. In this instance, the compromise which was adopted was one which ensured the operation of a practical and effective judicial institution.