

NORTHERN IRELAND LEGAL QUARTERLY

Post Devolution: Crystallising the Future for Regional
Government in England (Peter Leyland)

Disability, Degradation and Dignity: The Role of
Article 3 of the European Convention On Human
Rights (Anna Lawson)

The Significance of Corporate Collective Power to
New Labour's Project: Market Values, Corporate
Regulation and the Possibilities for a Progressive
Politics (Gary Wilson)

Conceptualising the Constructive Trust
(Andrew D. Hicks)

'Protestant' Political Theory and the Significance of
Rights (Sean Coyle)

The Forest-path of Law and the Wolf-belly of Justice:
Legal Theory and Fairy Tale (David Gurnham)

Psychiatric Harm and the Involuntary Participant:
"A Story of the Ebb and Flow of Tort Liability"
(Vivien Pickford)

Full table of contents inside

Vol. 56 No. 4

Winter 2005

CONTENTS

Post Devolution: Crystallising the Future for Regional Government in
England (*Peter Leyland*) 435

Disability, Degradation and Dignity: The Role of Article 3 of the
European Convention on Human Rights (*Anna Lawson*) 462

The Significance of Corporate Collective Power to New Labour’s
Project: Market Values, Corporate Regulation and the Possibilities for
a Progressive Politics (*Gary Wilson*) 492

Conceptualising the Constructive Trust (*Andrew D Hicks*) 521

‘Protestant’ Political Theory and the Significance of Rights
(*Sean Coyle*) 551

The Forest-path of Law and the Wolf-belly of Justice: Legal Theory and
Fairy Tale (*David Gurnham*) 585

Psychiatric Harm and the Involuntary Participant: “A Story of the Ebb
and Flow of Tort Liability” (*Vivien Pickford*) 602

BOOK REVIEWS

*Jurisprudence or Legal Science? A Debate about the Nature of Legal
Theory*: Sean Coyle and George Pavlakos 611

*A Star Chamber Court in Ireland: The Court of Castle Chamber,
1571-1641*: Dr. Jon G. Crawford 613

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

ISSN 0029-3105



POST DEVOLUTION: CRYSTALLISING THE FUTURE FOR REGIONAL GOVERNMENT IN ENGLAND

*Peter Leyland, Professor of Public Law, London Metropolitan
University**

Introduction

Devolution in Scotland, Wales, Northern Ireland has established a revised structure for territorial government in the United Kingdom. These arrangements have been one of the most far reaching re-allocations of powers in the recent history of the United Kingdom. At a time when government has become increasingly complex and multi-layered the devolution measures have drawn attention to the lack of an equivalent level of government in England. Although London has a directly elected Mayor and Assembly exercising a number of strategic London-wide functions, following the failure to gain popular approval in the referendum held in the North East region in November 2004 the prospects for directly elected regional government in the remainder of England have, in the short term at least, receded. However, it is clear that despite the lack of any equivalent form of elected government for the English regions, there have been significant steps towards introducing a form of regional governance throughout England. This article seeks to provide a critical discussion of the English regional governance initiative and to assess the prospects for the introduction of elected government for the English regions.

It is argued that any such trends in regional government and governance need to be understood in the light of what has been termed 'a multi-layered constitution'. In an institutional sense the United Kingdom has witnessed the introduction of what has been identified as multi-levelled (as opposed to layered) government. Looking first at the recognised institutional structure. At supra-national level the UK and other member states are subject to EU law and international obligations. At the level of the nation state, there are elected national parliaments and executives. At sub-national level, most European nations have a configuration of federal, regional/devolved government and below this also of local government. In addition to these formal levels, multi-layered constitutionalism recognises additional layers not only above and beneath, but between the formal institutional levels, and also sees the layers running horizontally. Taking for a moment the formation of geological strata as an analogy, the 'layers' have been laid down, and are being laid down as part of a dynamic process.¹ Further, the layers interact at

* I would like to thank Dr. Gordon Anthony and Professor George Jones for very helpful comments on earlier versions of this article. Many thanks also to Simon Crawshaw, Communications Officer of the North West Regional Assembly for his help with this research.

¹ The dynamic element is particularly evident in an uncodified constitution where statutes are passed and conventions constantly develop in relation to important aspects of the constitution.

horizontal as well as at vertical levels, with some layers more prominent than others. An important consequence of this modified state formation is that the relatively clear division between public and private elements has become obscured. The public sector has been shrinking with the emergence of what has been termed 'a contract state'. Governance, as opposed to government, is not necessarily delivered through formal mechanisms of civil service and local government but often by the private sector. The multi-layered constitution recognises: supra-national global considerations;² that the contemporary civil service has been organised into executive agencies at arms length from the core decision making; state contracting with the private sector and privatisation accompanied by statutory regulation;³ soft law agreements;⁴ informal remedies;⁵ human rights requirements;⁶ the emergence of e-government⁷ and so on. In the light of such far reaching transformation, it has been increasingly important to not only map the revised institutional framework of government and governance, but also to review the mechanisms of accountability that apply to it. In this article the case of devolution in Scotland, Wales and Northern Ireland and governance in London and the English regions provides an excellent example of both the different formal institutional levels and of multi-layered governance.

The approach will be to first look at the structural framework of devolution and the broad pattern established for the allocation of functions and powers between central and devolved government. Second, there will be discussion of the wider constitutional and administrative implications for England caused by devolution. Third, it will be explained how the Regional Development Agencies Act 1998, which was introduced concurrently with

² Including the intervention of multi-national corporate power and international obligations. See, e.g. D. Held, A. McGrew, D. Goldblatt, J. Perraton *Global Transformations: Politics, Economic, Culture*, London, Polity, 1999; W. Hutton *The World We're In*, London, Abacus, 2003.

³ See, e.g. Utilities Act 2000 and C. Graham *Regulating Public Utilities*, Oxford, Hart Publishing, 2000.

⁴ Informal concordats have been employed widely to assist with the implementation of devolution to Scotland, Wales and Northern Ireland. See, e.g. R Rawlings 'Concordats and the Constitution' 116 *Law Quarterly Review* (2000) 257.

⁵ The Citizen's Charter initiative has been rebranded under Labour as *Public Services Reform* and is now based on four guiding principles which regard the citizen as customer: national standards in public services; devolution to local leaders; more flexibility in delivery; greater customer choice. See P. Birkinshaw *Grievances, Remedies and the State*, 2nd ed., London, Sweet & Maxwell, 1995. Also for the wider trends toward informal dispute resolution see C. Glasser and C. Harlow 'Legal Services and Alternatives: The LSE Tradition' in R. Rawlings (ed.) *Law, Society and Economy*, Oxford, Oxford University Press, 1997.

⁶ The European Convention on Human Rights was effectively incorporated as a domestic 'Bill of Rights' by the Human Rights Act 1998.

⁷ Other commentators have developed 'Differentiated policy models' and 'policy networks' to discuss the substrata beneath the surface of the institutional levels of government and main stream legislative provisions but such analysis focuses on public policy rather than accountability. See, e.g. R. Rhodes *Understanding Governance: Policy Networks, Governance, Reflexivity and Accountability*, Buckingham, Open University Press, 1997; D. Richards 'New Labour, the State and the Constitution' in S. Ludlam and M. J. Smith (eds) *New Labour in Government*, Basingstoke, Macmillan, 2001.

devolution, cleared the path for a layer of devolved governance in the English regions.⁸ Fourth, there is consideration of the manner in which powers and functions which would have been re-distributed between central government and the English regions as part of the governments proposals for elected regional assemblies. The article concludes with an assessment of the way forward for English regional government considered as part of what is termed a 'multi-layered constitution'.

Constitutional Characteristics of Devolution

Although constitutional changes have been extensive under the Labour governments elected since 1997, the devolution arrangements were not undertaken as part of a wider strategy of constitutional transformation.⁹ Rather, the package of measures contained in each devolution statute has amounted to a distinct model for particular problems associated with the territorial entity. These devolution schemes were designed pragmatically. They have conferred sufficient power to satisfy local political aspirations while also addressing dissatisfaction felt at the periphery with the centralising tendencies emanating from Westminster. As a prelude to the discussion of the provisions intended for the English regions that follows, it will be helpful to draw attention to the main characteristics of each of the models of devolution which have been introduced to date.

Scottish Devolution

Although central government at Westminster has been careful to retain sovereignty Scotland has been granted the strongest form of devolution. Under the Scotland Act 1998 it has a single chamber Parliament of 129 elected members¹⁰ which meets for a 4 year term.¹¹ Following an election to the Scottish Parliament a government is formed after Parliament has nominated a First Minister. In turn, the First Minister is empowered to appoint ministers from MSPs to form a Scottish Executive. The executive is roughly equivalent to the Cabinet (under the Westminster system) and the ministerial appointments are made subject to Royal approval.¹² The Scottish Parliament and Executive have been given responsibility for many aspects of domestic policy within Scotland including: education, law, courts, prisons, judicial appointments, economic development, agriculture, fisheries, local government, the environment, housing, passenger and road transport, forestry and the arts. At the same time there are particular functions reserved for Westminster.¹³ The Scottish Executive is the administrative organ which

⁸ There has been some revision of the structure of local government and new initiatives which attempt to re-invigorate the existing system.

⁹ For more detailed discussions of UK devolution see N. Burrows *Devolution*, Sweet & Maxwell (2000); V. Bogdanor *Devolution in the United Kingdom*, Oxford University Press, (1999).

¹⁰ Additional member system of election. Scotland Act 1998 s.6 herein after referred to as the SA.

¹¹ SA s.2 an election can be called prematurely in certain circumstances.

¹² See SA s.47.

¹³ See SA 1998, S.29 and Sch.4.

has taken over most of the powers of the Scottish Office,¹⁴ and it is responsible for the implementation of policy in Scotland. The Scottish Parliament in addition to exercising an oversight function over the executive has been given the power to pass primary legislation.¹⁵ This power is strictly limited to matters under the scope of its legislative competence.¹⁶ Further, the Scottish Parliament and the executive has been granted tax raising powers (so far unused)¹⁷ but the method of funding remains mainly under the pre-existing Barnett formula¹⁸ determined by the Westminster government.¹⁹ Devolution has introduced a locally elected Parliament and given Scotland improved mechanisms of accountability that involve an increased level of political participation.²⁰

Welsh Devolution

Welsh devolution²¹ is based on a National Assembly for Wales elected every 4 years with 60 members which is required to form policy and take decisions in its particular areas of responsibility.²² A Welsh First Secretary and a cabinet style government is formed from the Assembly members following an election. The Welsh Assembly Government took over by transfer orders most of the administrative functions of the Secretary of State for Wales.²³

¹⁴ Prior to devolution the Secretary of State for Scotland was the Cabinet minister with executive responsibility for Scotland.

¹⁵ For a discussion of the SPs law making role see A. Page 'A Parliament that is Different? Law Making in the Scottish Parliament' in R. Hazell (ed.) *Devolution, Law Making and the Constitution*, Exeter, Imprint Academic, 2005.

¹⁶ The interpretation of the Scotland Act 1998 s.29(2)(b) is important constitutionally, since it provides that a matter is outside the competence of the Scottish Parliament if it relates to any matters reserved for the Westminster Parliament. If this section were to be given a narrow definition, it would restrict the Scottish Parliaments legislative capacity. The Judicial Committee of the Privy Council (similar in composition to the judicial committee of the House of Lords) is required to determine whether the Scottish Parliament enacts legislation which is outside its powers and it is empowered to set aside such legislation. See G. Gee 'Devolution and the Courts' in R Hazell and R. Rawlings (eds.) *Devolution, Law Making and the Constitution*, Exeter, Imprint Academic, 2005.

¹⁷ These powers have not so far been used.

¹⁸ Referred to as the Barnett Formula. See A. Midwinter, 'Territorial Resource Allocation in the UK: A Rejoinder on Needs Assessment' *Regional Studies*, Vol 36, 2002, 563-567. For a robust defence of the Barnett formula in preference to a needs based formula. It is argued that the added complication of reaching a formula for needs assessment would lead to greater scope for disagreement and more territorial lobbying.

¹⁹ See Scotland Act 1998, Part IV ss.73-79.

²⁰ An impressive feature in Scotland and Wales is the proportion of women in the Scottish Parliament and the Welsh Assembly.

²¹ For a comprehensive study of Welsh devolution see R. Rawlings *Delineating Wales, Constitutional, Legal and Administrative Aspects of National Devolution*, Cardiff, University of Wales Press, 2003 and R. Rawlings 'The New Model Wales' *Journal of Law and Society*, 1998 461-509. Section IV 'Lawmaking and Adjudication' deals with the functions of the assembly.

²² The main features of Welsh devolution are contained in the Government of Wales Act 1998, hereinafter referred to as the GWA.

²³ GWA s.22(2) and sch.2 e.g. agriculture, forestry, fisheries and food, environmental and cultural matters, economic and industrial development,

The Cabinet members have the equivalent of departmental responsibility for their given policy areas. But whereas the Scottish Parliament is granted general competence subject to the reserved matters under the Scotland Act, in the case of Wales powers are conferred in respect of particular areas of policy, and also for the Welsh quangos.²⁴ Accordingly, the Assembly has responsibility for conducting executive scrutiny over the Welsh Assembly Government and quangos. However, unlike the Scottish Parliament the Welsh Assembly does not have the power to pass primary legislation,²⁵ nor does it have any tax raising powers. The Welsh Assembly Government relies on the Secretary of State for Wales to sponsor bills required in Wales through the Westminster Parliament.²⁶ However, the Assembly is able to pass secondary legislation. As in Scotland the financial allocation for devolved government in Wales comes from Westminster, and it is based on the Barnett block grant formula. Although devolution in Wales has been generally welcomed as a success, there have been calls for modifications. The Richard Commission which was set up in 2002 to look into increasing the powers of the Welsh Assembly reported in March 2004. It recommended that an expanded Assembly elected by (STV) should be given full legislative competence on a similar basis to the Scottish Parliament and that there should be a split between the Assembly and the Welsh executive.²⁷ Following the May 2005 General Election it seems likely that some increased powers will be granted to the Welsh Assembly but that a referendum will be required before the Assembly is given full legislative powers.²⁸

Devolution in Northern Ireland

The main objective in Northern Ireland was to accommodate the deep-seated political differences which have existed between Unionist and Nationalist communities,²⁹ and, at the same time, establish a system of devolved

education and training, health, housing, local government, social services, sport and tourism, town and country planning, transport, water and flood defences and the Welsh language.

²⁴ This term refers to non-departmental governmental organisations, funded and appointed by government, *e.g.* Welsh Health authorities, Welsh Tourist Board.

²⁵ Since the introduction of devolution there have been calls to give the Welsh Assembly the power to pass laws but until such time as the legislation is amended, it is the responsibility of the Secretary of State for Wales to guide Welsh legislation through the Westminster Parliament. The Welsh Assembly has the power to pass secondary legislation.

²⁶ See K. Patchett 'Principle or Pragmatism? Legislating for Wales by Westminster and Whitehall' in R. Hazell and R. Rawlings (eds.) *Devolution, Law Making and the Constitution*, Exeter, Imprint Academic, 2005.

²⁷ See Report of the Commission on the Powers and Electoral Arrangements of the National Assembly of Wales (2004).

²⁸ The intention of introducing a Bill to increase the powers of the Welsh Assembly was announced in the Queen's speech May 2005 – see, R. Rawlings 'Hastening Slowly: The Next Phase of Welsh Devolution' 2005 *PL*, 824.

²⁹ To reconcile these differences at other levels: firstly, the Act establishes an Equality Commission for Northern Ireland to ensure that the equality of opportunity obligations under ss.75 and 76 of the NIA 1998 is complied with in all areas. Second, a Northern Ireland Human Rights Commission (NIHRC) was established to keep under review the adequacy and effectiveness in Northern

government to replace Stormont³⁰ which collapsed in the early 1970s. The political institutions in Northern Ireland have only operated sporadically due to political difficulties over issues that included decommissioning. First, the Northern Ireland Act 1998 provides for a directly-elected Northern Ireland Assembly, with law-making powers, consisting of 108 members elected every four years. These members are elected by Single Transferable Vote (STV) from 18 six-member constituencies. The Assembly has legislative powers over those matters which previously fell under the responsibility of the six Northern Ireland Government Departments namely, for agriculture, economic development, education, environment, finance and health and social services but these have been divided into ten ministerial portfolios post devolution. The Presiding Officer of the Assembly examines proposed legislation to ensure it falls within the legislative scope of the Assembly.³¹ Legislation passed by the Assembly requires the Royal Assent.³² In Northern Ireland there is a unique system of compulsory power sharing at every level of decision making to ensure joint participation by both communities in the processes of government.³³ In common with Wales, no tax raising powers are granted to the Northern Ireland Assembly.³⁴ In order to accommodate Nationalist aspirations for a united Ireland the system of government is linked to that of the Irish Republic. To satisfy Unionists fears that the Union could be severed without consent, there are East-West links within United Kingdom. The North-South Ministerial Council brings together members of the executive of the Northern Ireland Assembly and representatives of the Irish government for the purposes of co-operation on issues of common interest. The British-Irish Council is a body to consider broader mutual interests with the UK. It consists of representatives from the Scottish Parliament, Welsh Assembly, Channel Isles and Isle of Mann. Also, nationalist aspirations are safeguarded by a consent principle that ensures that Northern Ireland will remain a part of the United Kingdom only while a majority of the electorate wishes it to do so.³⁵

Ireland of the law and practice relating to the protection of human rights. The NIHRC has a balanced membership representing the local NI community.

³⁰ B. Hadfield 'The United Kingdom as a Territorial State' in V. Bogdanor (ed.) *The British Constitution in the Twentieth Century*, Oxford, Oxford University Press, 2003, see p.595ff for discussion of devolution and the Irish question.

³¹ The NIA s.5(6) further provides that this law making power should not affect the sovereignty of the UK Parliament. For a discussion of the law making powers of the NI Assembly see G. Anthony and J. Morison 'Here, There and (Maybe) Here Again: The Story of Law Making for Post-1998 Northern Ireland' in R Hazell and R. Rawlings (eds.) *Devolution, Law Making and the Constitution*, Exeter, Imprint Academic, 2005.

³² NIA s.5.

³³ Cross community support to ensure participation is defined under NIA s.4(5) and is required for, among other things, the election of First Minister and Deputy First Minister, Assembly's Presiding Officer, Assembly standing orders and approval of the annual budget.

³⁴ As for Scotland and Wales the funding for the Northern Ireland Executive is provided by central government block grant calculated under the Barnett formula.

³⁵ See s.1 Northern Ireland Act 1998.

Mayor and Assembly for London

The most significant change in the structural reorganisation of regional government in England since the election of the Labour government in 1997 has been the introduction of an elected London Mayor and Greater London Assembly. The restoration of a strategic level of government for London can be cited as further evidence of the Labour government's commitment to the devolution of power from the centre. The removal of the Greater London Council in 1985 by the Thatcher government had taken away a layer of government which had the task of co-ordinating the strategic aspects of policy and administration on a metropolitan basis. In dealing with such issues an all London body has a role complementary to that of the inner and outer London boroughs. The formation of the Greater London Authority (GLA) establishes an intermediate level of government directly accountable to the local electorate in London.

After a referendum approving the principle, the Greater London Authority Act 1999 led to the introduction of a Greater London Authority consisting of a Mayor and Assembly for London. In May 2000 and May 2004 elections were held in which Londoners were able to vote separately for a Mayor and for the Assembly which is elected by an additional member system of proportional representation. This additional member system which is similar to that used for elections to the Scottish Parliament and the Welsh Assembly was preferred, since it gives a more balanced party representation.³⁶ The Mayor and Assembly is responsible for spending an annual budget which on current figures will be approximately £9.6 billion for 2005-06.³⁷ This new level of London wide government also restores a form of democratic accountability for a range of important services and bodies.³⁸ The Mayor is placed at the head of the executive and is directly responsible for the strategies the GLA adopts to achieve its objectives. He is also responsible for the quality and effectiveness of the services which the authority delivers. The system is based on a clear separation of powers, with the Assembly holding the Mayor to account for his/her policies. In this capacity the Assembly is able to question the Mayor and the Mayor's staff by holding public hearings on issues of importance. In comparison with the Scottish, Welsh and Northern Irish devolution models the Mayor and the GLA have a much more limited range of responsibilities and funding powers, but the office extends to cover crucial strategic areas such as transport, planning, environment, crime and policing and sustainable development. But central government across the river at Westminster has been careful to retain

³⁶ Fourteen members represent the constituencies and eleven top up members represent London as a whole. See Greater London Authority Act 1999 s.2.

³⁷ See Published Budget 2005-2006; Mayor's Annual Report 2004-2005 at www.london.gov.uk/approot/mayor/annual_report/.

³⁸ The main areas of responsibility for the new authority include: transport (an integrated strategy for London and traffic management and regulation); economic development (responsibility for London Development Agency); Police, fire and emergency services (the Act creates a new Metropolitan Police Authority); Planning (requirement to develop a land use strategy for London); Environment (air quality and waste management); Culture (museums, library services and the arts).

ultimate control over certain aspects of high profile policy areas e.g., London transport.³⁹

Devolution and Governance

In one sense each devolution statute can be regarded as equivalent to a constitution for its respective part of the UK.⁴⁰ It will be explained in the next section that looked upon from a vertical axis devolution has constitutional implications for the inter participation of the actors representing the varying levels of government.⁴¹ Equally, it can be seen that rupturing the vertical relationship by creating a new level of devolved government has required realignment of administrative processes across the horizontal plane as well as the vertical axis. New strategies have been required to cope with the complex policy networks responsible for policy delivery as part of multi-layered governance. Looked at in terms of administration it can be said that to a certain extent devolution has put a seal of democratic legitimacy on administrative structures that were already largely in place. For example, Scotland has never been wholly incorporated into the UK and always retained its own distinctive church, legal and educational systems.⁴² All these policy areas were distinctive under the Scottish Office. To allow for such differences pre-devolution, Westminster legislation intended for Scotland (and Wales) followed a different procedure.⁴³ At the point of policy implementation, the growth of executive devolution was a feature of the system of government of the twentieth century.⁴⁴ Power in many policy fields had come to be exercised by officials under the direction of a Secretary of State for Scotland and for Wales. Indeed, for the most part, the powers that have been devolved in Scotland,

³⁹ There has been a clash with central government over the Public Private Partnership (PPP) method for regenerating London Underground imposed by central government. The weakness of the mayor's legal position was demonstrated in judicial review decisions. See *R v (1) London Underground Ltd (2) London Regional Transport, ex parte Transport for London* [2001] EWHC Admin 637; *R v Mayor of London, ex parte (1) Westminster City Council (2) Cathy Preece (3) Gareth Adamson*, Lawtel 31/72002. See P. Wintour 'Judgment day for London's Mayor' *The Guardian*, July 23 2002

⁴⁰ See J. Hopkins *Devolution in Context: Regional, Federal and Devolved Government in the European Union*, London, Cavendish, 2002 at p.168; B. Winetrobe 'Scottish Devolution: Aspirations and Reality in Multi-Layer Governance' in J. Jowell and D. Oliver (eds) *The Changing Constitution*, 5th ed., Oxford, Oxford University Press, 2004 p.173ff.

⁴¹ See the introduction to M. O'Neill (ed.) *Devolution and British Politics*, Harlow, Longman, 2004 at p.5 'Current devolution represents a significant shift in . . . historical arrangements'.

⁴² See, e.g. the Treaty of Union 1707. Article 18 guaranteed Scottish private law. Scottish Act of Security 1706 secured the protestant religion and the Presbyterian Church.

⁴³ The role of the Scottish/Welsh Grand Committee. See A. Bradley & K. Ewing *Constitutional and Administrative Law*, 12th ed., Essex, Longman, 1997 at p.201.

⁴⁴ Enforced compliance with central government policy was possible even where the local electorate opposed the policy. See D. Griffiths 'The Welsh Office and Welsh Autonomy' [1999] *Public Administration* 793-807 at p.805; I. Holliday 'Scottish Limits to Thatcherism' 1992, *Political Quarterly* 448-459.

Wales and Northern Ireland⁴⁵ have corresponded to the remit and responsibilities of the previous departments of state.

A fundamental problem is that the devolution legislation merely provides an outline which has been sketched out to varying degrees of detail to describe each of the respective systems of devolved government, in the main by specifying the principal institutions and the disparate powers conferred on each administration.⁴⁶ At an administrative level devolution needed to be incorporated into the existing uncodified constitution. Sub-rules, conventions and concordats have been adopted that exist and work outside the legislative framework. Inter-governmental relations have been addressed by informal processes. In particular, concordats has been drawn up between central government and the devolved administrations as a principal mechanism for the implementation of devolved government. Concordats have been referred to as a form of 'codification' of the processes of government⁴⁷ and it is widely acknowledged that the concordats are of practical importance for the administrative implementation of devolution⁴⁸. They have contributed to the continuity and smooth transition of policy and have also helped facilitate policy co-ordination between the many overlapping layers of modern governance.⁴⁹ That is to say a similar process of informal links and agreements is evident further down the policy chain and in relation to government regulation.⁵⁰ Concordats emerged fairly late in the process and 'represent a further step down the road of juridification in the

⁴⁵ In the case of Northern Ireland many of the devolved powers that had been vested in the Stormont system of government (which was suspended in 1972) reverted to the Northern Ireland Office. V. Bogdanor *Devolution in the United Kingdom*, Oxford, Oxford UP, 1999, p.99.

⁴⁶ Scotland Act 1998, e.g. Part I Parliament, Part II The Scottish Administration, Part III Financial Provisions, Part IV Tax Varying Powers. Sch.5 specifies reserved matters. Government of Wales Act 1998 Part I Assembly, Part II Assembly Functions, Part III Assembly Procedure, Part IV Assembly Finance, Part V other provisions about the Assembly, Part VI Reform of Welsh Public Bodies, Schedule 3 specifies functions to be transferred rather than 'reserved matters' for Westminster. Northern Ireland Act 1998 Part II Legislative Powers, Part III Executive Authorities, Part IV The Northern Ireland Assembly; Part V North-South Ministerial Council and British Irish Council, Part VI Financial Provisions, Part VII Human Rights and Equal Opportunities. Some commentators regard each statute as equivalent to a local constitution.

⁴⁷ R. Rawlings 2001 'The Shock of the New: Devolution in the United Kingdom' in E. Riedel (Hrsg.) *Aufgabenverteilung und Finanzregimes im Verhältnis zwischen dem Zentralstaat und seinen Untereinheiten*, Baden Baden, Nomos Verlagsgesellschaft, 2000 p.83.

⁴⁸ The Department of Constitutional Affairs states that: 'The Concordats are intended to promote constructive co-operation and communication. They set out a working framework within which Ministers and officials may continue and develop relationships between the administrations. The primary aim is to ensure that if either is planning action impinging on the responsibilities of the other, it gives adequate forewarning.'

⁴⁹ R. Rawlings 'Concordats of the Constitution' [2000] *LQR* 256-286 at p.276.

⁵⁰ P. Leyland 'Utility Regulation in an Age of Governance' in N. Bamforth and P. Leyland *Public Law in a Multi-Layered Constitution*, Oxford, Hart Publishing, 2003.

form of “bureaucratic law”⁵¹ Thus for the picture to come into any sort of focus we find that, certainly at an administrative level, reference has to be made to a developing body of *soft* law. A notable feature of the hand over of power since devolution has been a smooth transition with little or no disruption to the delivery of policy.⁵² In many policy areas the operationalising of the devolved bodies has allowed funding to be allocated according to local priorities.

In sum, the devolution arrangements have provided a new level of government subordinate to Westminster on an asymmetrical basis which, not only provides the electorate in each nation with the right to vote on the basis of proportional representation, but also introduces a unique set of democratic institutions and processes for Scotland, Wales, Northern Ireland and London. The Scottish Parliament has been granted limited powers to raise taxes, but otherwise the financial formula that determined the overall sums available to Scotland, Wales and Northern Ireland has been left untouched.

Constitutional Implications of Devolution

Viewed from a constitutional standpoint devolution has produced an inequality of political representation, an issue sometimes referred to as the ‘West Lothian question’. As one commentator has put it: ‘It raises serious questions about the role of MPs as members of the UK parliament, and about the nature of the Union itself. The Union has traditionally been built on an equality whereby all members can vote on all matters, regardless of territorial extent of their application, as members of a single parliamentary body.’⁵³ If, for instance, we look at the current position in Scotland, the introduction of a Scottish Parliament and Executive with considerable power plays havoc with the notion of representative government in the United Kingdom. MPs representing English, Welsh and Northern Irish constituencies are no longer able to vote on devolved matters in Scotland but Scottish MPs at Westminster retain the right to vote on domestic policy for the rest of the UK.⁵⁴ Further still, by the transfer of many domestic functions to the Scottish Parliament Scottish Westminster MPs have a greatly reduced role to play in relation to their constituents. The obvious line of accountability for the devolved areas of domestic policy in Scotland is through their Scottish representatives (MSPs).

⁵¹ Rawlings 2000 at p.258. This article presents an in depth critical evaluation of concordats. See also M. Laffin, A. Thomas and A. Webb ‘Intergovernmental Relations After Devolution’ [2000] *Political Quarterly*, [2000] 223-245.

⁵² This hand over of power was made possible through a series of concordats drawn up between Whitehall and the new executives e.g. *Concordat between the Cabinet Office and the Scottish Administration*. The significance of concordats is discussed in greater detail below.

⁵³ See M. Russell & G. Lodge, ‘The Government of England by Westminster’, in R. Hazell (ed.) *The English Question*, Manchester University Press, 2005 (forthcoming).

⁵⁴ See R. Johnston, C. Pattie and D. Rossiter ‘Devolution and Equality of Representation in the United Kingdom: A Constitutional Mess’, *Political Quarterly*, Vol. 73, No. 2, April-June 2002 158-172 who also discuss the over representation of Scotland and Wales in the House of Commons.

Many aspects of policy can be introduced throughout England by central government with the approval of MPs representing Scottish, Welsh and Northern Irish constituencies without the government having to gain approval from English MPs alone. By way of contrast, a separately elected Parliament or Assembly decides equivalent matters in the devolved parts of the UK. In other words, an effect of devolution has been to throw open the question of parliamentary representation for England which in terms of political representation is treated less favourably to Scotland, Wales, Northern Ireland and London. How to address this question and the provision of an equivalent level of government for England is the focus of the remainder of this article⁵⁵.

Voices of discontent at the lack of provision for England have been heard most stridently in the socially disadvantaged English regions bordering Scotland and Wales that have missed out economically to the greatest extent from devolution,⁵⁶ but until recently there has been minimal attention to assessing the likely impact of these arrangements at the heart of the UK.⁵⁷ The Campaign for an English Parliament, as its name implies, presents a case for the introduction of an English Parliament. A proposal for an equivalent body to the Scottish Parliament might appear to have some merit since setting up a Parliament for England could provide the constitutional basis for correcting the glaring asymmetries relating to representation, accountability and administration which, as we have just noted, have been raised by devolution. However, there are obvious reasons why such a proposal has little chance of being implemented.⁵⁸ First, it has hardly any support within any of the mainstream political parties, or more widely with the English electorate. Second, a Parliament for England with equivalent powers to the Scottish Parliament would be a strong competitor to the Westminster Parliament which would no longer have a pivotal role in relation to domestic issues. Third, an English Parliament would be an additional elected political body requiring a fresh cohort of politicians at a time when there is already evidence of voter fatigue with existing elected political institutions. Fourth, establishing an English Parliament would prompt a debate about the financial provisions on which devolution is based and, in so doing, open up fresh controversy over the allocation of resources within the UK. In turn, such controversy might provide further impetus to the devolution process by prompting wider calls for Scottish independence. Fifth, an English Parliament would be very costly to introduce. It must be stressed that there has been limited popular support in England as a whole for an English Parliament, or indeed, for any comparable constitutional reform.

An alternative response to the problem pertaining to the voting rights of Westminster MPs, which is much less radical than resorting to an English

⁵⁵ The Regional Development Agencies Act 1998 introduced a new policy for economic regeneration by dividing England into 9 regions.

⁵⁶ See, *e.g.* Campaign for the English Regions: <http://www.cfer.org.uk/home.htm>; Giving the People a Voice: A Campaign for Yorkshire's White Paper (September 2001). Campaign for an English Parliament.

⁵⁷ V. Bogdanor, *Devolution in the United Kingdom*, Oxford, Oxford UP, 1999 p.266.

⁵⁸ The issue of an English Parliament is discussed in B. Hadfield 'The English and Welsh Questions' in J. Jowell and D. Oliver (eds.) *The Changing Constitution*, Oxford, Oxford UP, 2004 at p.246ff.

Parliament, would be to agree on a new convention or implement new procedures and rules within Parliament which would prevent Scottish and Northern Irish Westminster MPs from voting on legislation not applying in Scotland and Northern Ireland.⁵⁹ It should be pointed out that there is an obvious political difficulty with reaching such a resolution at Westminster in the immediate future. A Labour government which is heavily dependent on the support of Scottish and Welsh Labour MPs to get its legislation through Parliament is unlikely to agree to a change which has the effect of disabling its political power.⁶⁰ On the other hand, it has suited the Conservative Party in opposition to take up the call for 'English votes for English laws'.⁶¹ A change in the rules which prevented MPs outside England from voting on English bills would effectively mean that Labour would no longer have an overall majority in the House of Commons for votes on legislation which concerned England alone and such a change would be a means of consolidating the power base of the Conservative Party in England where its support is normally at its strongest.⁶² In a different sense it has been pointed out that recent proposals to enhance the law making capacity of the Welsh Assembly⁶³ are likely to have an incidental impact on England. It will mean that as the revised and distinct procedure for law making in Wales is introduced: '...distinctively "English" legislation in the Westminster Parliament is on the horizon, a not insignificant innovation driven here from the other side of Offa's Dyke'.⁶⁴

A related question which would also need to be addressed concerns the role of the House of Lords, or reformed second chamber, in dealing with legislation directed at particular parts of the United Kingdom. Life peers and the remaining hereditary peers originating from any part of the nation are presently eligible to sit and vote in the House of Lords. How could a rule restricting voting rights in the House of Commons be reconciled with the practice and procedure in a reformed House of Lords? Should a reformed second chamber have a regional element to its composition?⁶⁵ There are no

⁵⁹ See M. Keating 'The UK as a post-sovereign polity' in M. O'Neill (ed.) *Devolution and British Politics*, Harlow, Longman, 2004, p.323.

⁶⁰ Russell and Lodge 2005 above.

⁶¹ Both William Hague and Michael Howard as opposition leaders have supported this idea.

⁶² A further problem in introducing any restrictive rule over the way legislation is considered by MPs as it passes through Parliament concerns the technical difficulties in drafting legislation. *E.g.* where there are mixed clauses some of which only apply to particular parts of the UK. For detailed discussion of the implications of changing voting rules see: B. Hadfield 'Devolution, Westminster and the English Question' *PL* [2005] p.301.

⁶³ See *Better Governance for Wales*, Cm.6582, 2005. The proposed change will be in phases. First, a more consistent approach to legislative drafting. This has already been announced; second, to combat the legislative bottleneck at Westminster a new form of Order-in-Council for Wales has been devised; and the third phase, would grant full law-making powers to the Welsh Assembly, but only after approval by referendum triggered by a two thirds support in the Assembly.

⁶⁴ R. Rawlings 'Hastening Slowly: The Next Phase of Welsh Devolution' [2005] *PL* 824-852 p.841.

⁶⁵ See, *e.g.* M. Russell and R. Comes 'The Royal Commission on Reform of the House of Lords: A House for the Future?' [2001] 64 *MLR* 82-99 at p.90 and Royal

firm proposals to modify the procedure or composition of the Westminster Parliament in accordance with the options discussed. However, the speculation over possible further constitutional change illustrates the dynamics of devolution. It has prompted debate over further modifications to the constitution to address the issues it raises.

Governance for the English Regions

Since coming to office in 1997 the Labour government has repeatedly declared its intention not only to ‘modernise government’ but also to ‘bring democracy closer to the people’.⁶⁶ However, comprehensive reform through the introduction of some form of devolved government for England and its regions, which is by far the greater part of the UK in population terms with over 50 million people, was a notable omission from the devolution arrangements for the remainder of the UK set up between 1998 and 1999.⁶⁷ In part, the explanation for the absence of constitutional modifications relating to England is because England has only ever existed as a ‘state of mind’ not as a ‘consciously organised political institution’.⁶⁸ Put more precisely: ‘It is the supposedly unified and homogeneous nature of England which has in large part been responsible for the unitary state’.⁶⁹ England has tended to share aspects of its governance with Wales and to a lesser extent Scotland and Northern Ireland. Moreover, in discussing the West Lothian question we have already encountered the difficulties in overcoming the asymmetry caused by devolution without at the same time disrupting the balance of the constitution. Nevertheless, campaign organisations seeking elected regional government in some of the English regions outside London have been somewhat more successful in making their case.⁷⁰ The Labour government has introduced a layer of English regional governance in preparation for what it promised would be a more ambitious programme of elected English regional government.

Viewed from a governance perspective, the issue of policy co-ordination at regional level in England has been recognised for some time.

“The hollowing out of the departmental functions to a plethora of agencies, non-departmental public bodies and private sector entities has placed a new emphasis on the need for government organizations to engage in intergovernmental diplomacy and coalition-building in order to deliver their own priorities. An unexpected outcome of the atomizing of government has been the corresponding recognition by officials of the need to

Commission on Reform of the House of Lords, *A House for the Future*, Cm 4534 (2000) (Wakeham Commission).

⁶⁶ See, e.g. *Modern Local Government: In Touch with the People* (1998) Cm 4014.

⁶⁷ J Tomaney “New Labour and the English Question”, *Political Quarterly*, 1999 75-82. The possible introduction of an English Parliament and the prospects of regional government are discussed in the light of devolution for Scotland, Wales and Northern Ireland.

⁶⁸ R. Rose *Understanding the United Kingdom*, Harlow, Longman 1982, p.29.

⁶⁹ V. Bogdanor *Devolution in the United Kingdom*, Oxford, Oxford University Press, 1999 p.266.

⁷⁰ See, e.g. Campaign for the English Regions: <http://www.cfer.org.uk/home.htm>.

counter its effects through networking and their willingness to engage in diplomacy.⁷¹

In the early 1990's the Conservative government created the Government Offices of the Regions (ROs). The idea was to establish an integrated approach to policy making by bringing together government and its client groups. At this point as a result of the highly complex interweaving of different departments and diverse bodies covering different, but often related, policy areas, there was inadequate consistency in the operation of the wide variety of policy networks.⁷² Subsequently, some consolidation was achieved with the combining of departmental regional offices for the regions representing key departments under one roof.⁷³ It has been observed that 'the partnering region' relates to the novel forms of local democratic governance, including appropriate models of regional economic development, as well as the tools and practices of administrative and political cooperation across the various layers of government.⁷⁴ The regional policy of successive governments has been largely driven by a desire to co-ordinate policy at a regional level and equally a commitment to encourage economic development. Some commentators have not been convinced by the motivation behind such a policy, noting that the regional government initiative amounts to little more than 'another form of functional decentralisation'.⁷⁵

Under the 1997 Labour government the next step in the direction of regional government exactly coincided with devolution.⁷⁶ The Regional Development Agencies Act 1998⁷⁷ was a vehicle for the introduction of a new policy for economic regeneration.⁷⁸ The 1998 Act divided England into 8 regions (plus London) and created Regional Development Agencies (RDAs) in order to co-ordinate regional economic development and the new agencies were made responsible for formulating strategies for regeneration to promote business efficiency, investment and competitiveness in their areas⁷⁹ by forging links and formalising relationships with the business sector at regional level. The RDA's are mainly funded from central government and they are given

⁷¹ R. Rhodes, P. Carmichael, J. McMillan, A. Massey *Decentralizing the Civil Service: From unitary state to differentiated polity in the United Kingdom*, Buckingham, Open University, 2003 at p.147.

⁷² Stevens 2004 p.256.

⁷³ E.g. the Department of Trade and Industry, the Department of Employment, the Department of the Environment and the Department of Transport.

⁷⁴ R. Rawlings *Delineating Wales: Constitutional, Legal and Administrative Aspects of Welsh Devolution*, University of Wales Press 2004, p.433.

⁷⁵ R. Rhodes, P. Carmichael, J. McMillan, A. Massey *Decentralizing the Civil Service: From unitary state to differentiated polity in the United Kingdom*, Buckingham, Open University, 2003 at p.137.

⁷⁶ See J. Tomaney 'The Evolution of Regionalism in England' *Regional Studies*, Vol. 36.7 pp.721-731.

⁷⁷ R Hazell and B O'Leary 'A Rolling Programme of Devolution' in *Constitutional Futures: A History of the Next Ten Years*, Oxford, OUP, 1999, pp.36-40.

⁷⁸ See Regional Development Agencies Act 1998, Sch. I: East Midlands, Eastern, North East, North West, South East, South West, West Midlands, Yorkshire and the Humber and London.

⁷⁹ Regional Development Agencies Act 1998 ss.4 & 5.

limited powers to achieve this aim.⁸⁰ Financial allocations to the RDA's from the contributing central government departments⁸¹ are pooled into a single budget.⁸² The sum each RDA receives from this single pot is largely determined by a formula reflecting the economic situation in each region, which has been agreed between DTI and the RDAs. The formula uses a series of indicators as a proxy for the needs of each region, including the needs of businesses in the region. The most up to date data is used and a system of weightings is applied to the indicators to take account of existing funding commitments made by the RDAs. The North West is allocated £382 million, while the North East is granted £240 million. As we will shortly observe, equivalent bodies in Wales and, in particular, the Welsh Development Agency, were placed under the Welsh Assembly government, providing a direct element of democratic control and accountability.⁸³ Whereas the boards of RDAs are appointed by ministers and are only indirectly accountable to Parliament through the minister.

Another important reason for achieving a high degree of regional policy co-ordination has been to put the English regions in a position to bid for available EU funding which is allocated on a regional basis.⁸⁴ An extra tier of government opens up the opportunity of attracting more of the European Unions structural funds designated for the regions.⁸⁵ The EU dimension has encouraged the emergence of a greater degree of autonomous regional self government in Europe.⁸⁶ Spain has also been confronted with a combination of nationalist demands for self-government and economic pressures for greater autonomy, and Spain too has opted for a staged programme of asymmetrical devolution.⁸⁷

⁸⁰ Under s.5(2) of the Regional Agencies Act 1998 the agencies can give financial assistance, dispose of land or form or acquire and interest in a body corporate but only with the consent of the Secretary of State.

⁸¹ The departments are: Trade and Industry (DTI), Deputy Prime Minister (ODPM) formerly Environment, Education and Skills (DfES), Environment, Food and Rural Affairs (DEFRA) and Culture, Media and Sport (DCMS).

⁸² The total allocation for the year 2004-2005 was £1850 million. The amount for 2005-2006 has been set at £2163 million. The ODPM provides the largest contribution amounting to £1.5 but a substantial contribution from Trade and Industry £233 in 2004 - 2005. The contribution from the DTI rises to £463 million in £2005-2006.

⁸³ The government in Wales duly decided to abolish and modify some of the quangos it inherited.

⁸⁴ The White Paper notes at 8.16 that: 'Individual local authorities and regional bodies such as Regional Development Agencies have realised that, to make an impact in Brussels, they need to work together at a regional level to gain critical mass. The efforts of the UK's devolved administrations have demonstrated the benefits of such co-ordinated action'.

⁸⁵ M. O'Neill 'Great Britain: From Dicey to Devolution' *Parliamentary Affairs*, 2000, 69-96 at p.91.

⁸⁶ White Paper 2002 4.3 a regional assembly will take over the role currently performed by Government Offices on structural funds. The White Paper further states at 8.19 'Structural Funds have been the catalyst for strengthened links between the regions and the EU, and are one of the most visible signs on the ground of the benefits of EU membership.'

⁸⁷ See J. Hopkins *Devolution in Context: Regional, Federal and Devolved Government in the European Union*, London, Cavendish, 2002 at p.120ff.

Not only were RDA's introduced to develop and deliver economic strategies to secure better and more sustainable economic performance for their regions but, in addition, the Labour government established Regional Assemblies in all of the English regions.⁸⁸ The Assemblies were formed partly from politicians representing the various local authorities in the regions⁸⁹ (approximately two thirds); and partly from stakeholders in the form of economic and social partner organisations (approximately one third).⁹⁰ All members were nominated rather than being directly elected, and the role of the Assemblies has been to contribute to regional economic strategies and plans, and to scrutinise the delivery of such strategies and plans. The Assemblies vary in size from 117 to 36 members. For example, the North West Regional Assembly represents the North West region of England which includes the counties of Cheshire, Cumbria, Greater Manchester, Lancashire and Merseyside. The region has a combined population of 7 million and an annual budget of £2.3 million.⁹¹ The Assembly is part funded from local authorities. The remainder is allocated following a formal bidding process from strengthening Regional Accountability (SRA) and English Regions Network (ERN). The Assembly is staffed by full time officials responsible for providing support to Assembly members. The principal objectives of the North West Region Assembly are described in its constitution.⁹²

The Assembly promotes the interests of the region nationally and internationally and with European and Government agencies.⁹³ More specifically, it seeks to promote through joint action with the partnership of

⁸⁸ See Regional Development Agencies Act 1998 s.8 (1) If the Secretary of State is of the opinion-

(a) that there is a body which is representative of those in a regional development agency's area with an interest in its work, and

(b) that the body is suitable to be given the role of regional chamber for the agency,

he may by directions to the agency designate the body as the regional chamber for the agency.

⁸⁹ In some rural areas two tier County Councils and District Councils have been replaced by 46 unitary authorities. There are 36 single tier metropolitan borough councils in cities and towns and 32 London Borough Councils.

⁹⁰ Stakeholder is not defined in the White Paper, but the term refers to business, trade unions, the voluntary and community sectors, local authorities and other key partners in the public sector. Also referred to as the business community, social and environmental partners, and elected representatives from other tiers of governance.

⁹¹ The other Assemblies for which figures are readily available: Yorkshire and Humber Assembly £4.7 million; South East Region Assembly £4 million; South West Assembly £1.4 million; East of England Assembly £1.4 million.

⁹² Their work involves '... commissioning research, developing policies, liaising with partner organisations, producing documentation, lobbying and working to raise awareness of current issues in the region in relation to education, planning, health, transport, culture, communities, economic development, housing, European issues and the environment'.

⁹³ According to its own publicity published on its website: 'The North East Assembly has become a leading player in the region with a 'voice' on every important issue affecting the region. We are speaking for the North East at a regional, national and European level. Our role in representation, influence and lobbying extends to a whole range of areas that affect everyday lives'.

interests, the sustainable economic, social and environmental well being of the region. To reach this objective it must assess 'the strategic needs of the region so as to make a strong input into the Regional Strategy and scrutinise its preparation or review by the North West Development Agency (NWDA)'.⁹⁴ The Assembly was intended to provide a strong authoritative voice for the region and, at the same time, to provide a consultative framework between partners on how to meet the challenges facing the region. The Assembly acts as a kind of proto-decision making body. It then exercises an oversight function in respect of monitoring the NWDA's implementation of the strategies that have been agreed by the Regional Assembly. It must ensure that the developing agenda of bodies such as Regional Housing Board and the emergent Regional Transport Board remain fully accountable to the full Assembly and the region.

The Regional Assembly debates and determines the broad economic strategies and priorities for the region. It follows from this that the RDA and other public, and publicly funded bodies in the region, fully consider these strategies, indeed they are required to do so as part of their statutory responsibilities. The Regional Assembly also provides feedback to central government on whether reports and plans published by the regional development agency have the support of the regional assembly.

The potential for policy co-ordination in the regions by bringing together the various public and private interests as part of a governance narrative in what we have called here a 'multi-layered constitution' can be illustrated in respect to planning policy.⁹⁵ The Regional Assemblies as the regional planning body have been given a leading role in the co-ordination of planning policy as part of the Regional Spatial Strategy (RSS).⁹⁶ In theory it might be claimed that: 'Policy decision emerges as a consequence of activity within this network rather than through Parliament, and policy coordination is a by-product of inter-organisational power bargaining rather than a strategic activity'.⁹⁷ Under the latest central government initiative, the Regional Spatial Strategy,⁹⁸ the Assembly is required to link its economic, housing, transport and planning goals together in what is referred to as a broad 'spatial strategy'. Planning policy must focus on the needs of the region as a whole, but any policy that is drawn up must also highlight those areas that need more specific guidance or a different approach. This approach is designed to not only improve the co-ordination and the delivery

⁹⁴ See North West Assembly website.

⁹⁵ 'The Planning and Compulsory Purchase Act 2004 marked a significant change, formalising the regional planning arrangements that have existed in England for some 15 years'. M. Sandford and P. Hetherington 'The Regions at the Crossroads: The Future for Sub-National Government in England' in A. Trench (ed.) *The Dynamics of Devolution: The State of the Nations 2005*, Exeter, Imprint-Academic, 2005 p.92.

⁹⁶ Government Guidance, RPG 1.

⁹⁷ C. Stephens 'English Regional Government' in M. O'Neill (ed.) *Devolution and British Politics*, Harlow, Longman, 2004 at p.261.

⁹⁸ Planning and Compulsory Purchase Act 2004 s.1. The main difference between the Regional Spatial Strategy and the previous Regional Planning Guidance (RPG) is that the RSS is statutory and it will become part of the new style development plans called Local Development Documents (LDDs).

of regional policy, but also to achieve sustainable development within the region. The RSS is set to provide the overarching framework for growth in each of the regions over the next 15 to 20 years.

In conformity with statutory guidelines any such policy is required to embody more inclusive and participatory decision making.⁹⁹ The Assembly has to formulate corporate plans which must both demonstrate added value and relate to those issues which are relevant to the local authorities which they represent and also to their economic and social partners. But although the Assembly has a central role in regard to the planning process at regional level, it is crucial to remember that under the relevant statute ministerial direction provides a constant guide in the framing of the overall policy.¹⁰⁰ It has been argued that the combination of decentralisation with local participation, coupled with a strong element of ministerial direction, has emerged as a blueprint for ensuring effective compliance with policies from the centre.¹⁰¹ In regard to the input of central government, which is still able to intervene on the vertical axis to ultimately determine policy.

At this point a strong contrast can be drawn with all of the devolution models which grant executive power to the periphery over the functions that have been devolved. In particular, quangos are placed under the control of the devolved government. This establishes a fresh line of accountability to an elected body. It also means that the devolved government is free to abolish or re-shape many of the quangos under its remit, more or less at will. For example, the Welsh Assembly Government adopted a policy to only retain quangos where they are the most appropriate and cost effective means of carrying out the functions concerned.¹⁰² In the period following the introduction of devolution the Welsh Assembly Government had to reach some sort of accommodation with the Welsh Development Agency which was regarded by some Assembly members as an unwelcome competitor in managing the Welsh economy. A new relationship has been forged between the two bodies and as Professor Rawlings notes: 'The WDA is both a great survivor, even a direct beneficiary of the devolutionary scheme, and epitomizes the extended accountability of public bodies in Wales associated with the devolutionary development'.¹⁰³

As part of the multi-layered constitution there is evidence also of horizontal cross regional integration. The English Regions Network (ERN) has been formed as an umbrella organisation for the Regional Assemblies. The role of

⁹⁹ See Planning and Compulsory Purchase Act 2004 s.6 (1) which provides: 'For the purposes of the exercise of its functions under s.5, the RPB must prepare and publish a statement of its policies as to the involvement of persons who appear to the RPB to have an interest in the exercise of those functions.'

The websites of each of the Regional Assemblies sets out the broad characteristics of their RSS which are all closely related.

¹⁰⁰ Under s.1(2) of the Planning and Compulsory Purchase Act 2004 'The RSS must set out the Secretary of State's policies (however expressed) in relation to the development and use of land within the region'.

¹⁰¹ R. Rhodes, P. Carmichael, J. McMillan, A. Massey *Decentralizing the Civil Service: From unitary state to differentiated polity in the United Kingdom*, Buckingham, Open University, 2003 at p.137.

¹⁰² See Rawlings 2003 p.358.

¹⁰³ Rawlings 2003 p.363.

the ERN is first to bring members and officials together to discuss matters of mutual interest. Second, it has a role in co-ordinating the activities of the assemblies. For example, information can be shared in order to influence the development of policy and practice and the range of initiatives under way in the different regions can be formulated into an integrated framework. The Network also seeks opportunities for regional assemblies to strengthen the voice of the regions in Whitehall and Westminster. Transportation is a policy area with obvious cross regional implications.

The Proposed Elected Assemblies for the English Regions: What went wrong?

The RDAs and Regional Assemblies were slotted into place during Labour's first term 1997-2001 and these institutions were to form the basis for the next stage of devolution for the English regions. The Labour Party went into the 2001 general election claiming: 'there is no case for threatening the unity of the UK with an English Parliament or the denial of voting rights to Scottish, Welsh and Northern Ireland's MPs at Westminster'.¹⁰⁴ However, it recognised a democratic deficit in England, and it announced plans for a new level of elected government for the English regions. While the government made no general commitment to elected regional government for the country as a whole, it made a pledge to introduce directly-elected regional government in regions where the proposals were supported by a referendum.¹⁰⁵ The 2001 manifesto contained a promise to enhance the scrutiny functions of regional chambers as they took on more powers. As mentioned earlier, there has been little evidence of popular support among the electorate for more devolved government in England.¹⁰⁶ Moreover, it is very much open to question whether the White Paper proposals would have made a significant contribution in the nature and quality of government for the English regions. In assessing these proposals it is worth remembering that, even in its weakest form, devolution has been characterised by the following features: first, it extends to an established territorial entity. Second, devolution provides a form of elected Parliament or Assembly with powers over a form of primary legislation or secondary legislation; third, a wide range of functions were placed under the direction of the devolved government. Fourth, the elected assembly exercises an oversight function over the executive functions performed at devolved level. Fifth, devolution resulted in the bureaucratic apparatus of government in the form of the civil service and relevant non-departmental public bodies coming under the direct ambit of the devolved form of government; sixth, each type of devolved government was granted considerable discretion in how it spent the funds made available to it from central government. In other words, although the funding for devolution mainly derives from central government under block grant, according to the Barnett formula, the effect of having a pre-determined

¹⁰⁴ *Ambitions for Britain*, Labour Party Election Manifesto 2001, p.35.

¹⁰⁵ *Ambitions for Britain*, Labour Party Manifesto, 2001, pp.34-35.

¹⁰⁶ It has been pointed out that regionalism is not sufficiently embedded and is therefore only government backed official regionalism. See C. Stevens 'English Regional Government' in M. O'Neill *Devolution and British Politics*, Harlow, Longman, 2004, at p.253.

formula has guaranteed the broad limits of the funding allocation conferred from Westminster which is kept at arms length where funding is concerned.

A vote in a referendum in favour of regional government would have resulted in a region having an assembly elected by proportional representation.¹⁰⁷ The elected assembly would have consisted of between 25 and 35 members and the new elected body would have replaced the existing regional assembly. After it had been elected each assembly would have voted for an executive of no more than 6 members, with the remaining members performing an executive oversight function through scrutiny committees. Set against the devolved systems the proposed regional assemblies were to be given control over very limited policy domains. As well as the functions in the hands of the already constituted regional assemblies, the additional areas which might have been included were: economic development, housing, public health improvement and culture (currently under county councils and district councils). Social services, education and other local services would have remained with existing local authorities.¹⁰⁸

The government claimed that these arrangements were an attempt to improve policy delivery by ensuring better co-ordinated government at regional level. These bodies were to be granted limited powers. The elected regional assemblies would have been able to formulate *strategies* at three distinct levels: high-level targets; strategies for specific subject areas; and an overarching regional strategy.¹⁰⁹ However, it was not proposed to give the elected regional assemblies law making powers or powers over delegated legislation on a comparable basis to the powers granted in Scotland, Wales or Northern Ireland. RDAs were given the role of surrogate regional executives, retaining the task of developing a regional plan, although the elected assemblies would have been in a position to direct that changes be made to a plan before it was published. RDA's were to have been made accountable to the Assembly for their other activities, but would have otherwise retained their operational independence.

The main objectives of an elected Regional Assembly would have been to set out a number of high-level targets, which it would have been required to agree with central government. Assemblies were to be placed under a duty to promote sustainable development and improve the quality of life for their region, and each regional assembly was to have a co-ordinating responsibility for 'joining up' policy, while at the same time, it would have been required to develop strategies for strengthening the region. The prime function of Assemblies with the increased functions and powers contained in

¹⁰⁷ See White Paper: *Your Region, Your Choice: Revitalising the English Regions*, 2002, Cm 5511 The voting system for elected regional assemblies was to be the Additional Member System (AMS). The method already used for the Scottish Parliament, the Welsh Assembly and the Greater London Authority. It ensures the overall composition of an assembly will broadly reflect the votes cast for the different parties at the assembly election.

¹⁰⁸ The functions transferred to the assembly were not precisely defined and could vary from region to region, greatly adding to the asymmetry of any devolution of power.

¹⁰⁹ White Paper, above chap.4. Strategies will consist of detailed plans setting out how the high level targets can be reached.

the draft bill published in 2004 would be: 'to promote industry and develop the skills of the workforce according to regional priorities . . . [to] invest in the physical regeneration of housing and make related transport investments. In short, assemblies could assume the role of leading bodies in promoting regional regeneration'.¹¹⁰

The proposed system confirmed the special recognition for regional stakeholders. The White Paper suggested that a clearer decision-making framework would be achieved by giving the stakeholders better access to decision making processes. In common with the existing regional assemblies, the regional executive would have been required to ensure that relevant interests, both public and private, would be engaged in developing and delivering these strategies. The fact that assemblies would have been conferred with executive functions and powers to implement these regional policies represented yet another attempt by the government to bring decision making closer to the electorate, while at the same time identifying the involvement of stakeholders in the policy process.¹¹¹

According to the government the elected regional assemblies were not intended to add an extra tier of government,¹¹² nor would they result in the introduction of additional bureaucracy.¹¹³ The main reason the assemblies would not have duplicated the work of either the UK Parliament, or existing local authorities, was because the majority of their functions would have come from central government bodies, primarily the regional development agencies and the regional assemblies. This feature allowed the regional government initiative to be presented as representing a significant decentralisation of power. Secondly, the avoidance of duplication would have been achieved because, before a referendum was called on the desirability of an elected Regional Assembly, a review would take place to ascertain the best method for a unitary structure of local government for that region.¹¹⁴ The package on the table for democratic regional government would have allowed voters when making their choice in any referendum aware of the implications for the re-organisation of local government in their region.

The White Paper indicated that the elected regional assemblies were to receive a block grant from central government. Therefore, in practice, most of the funding would have come through a single Government payment. In common with other devolved government in the UK, an elected regional assembly would then have been in a position to decide how to use its allocation to address regional priorities. This discretion would have ranged over a much narrower set of policy domains. Subject to certain strict conditions, regional assemblies were to be given power to borrow money on

¹¹⁰ J. Tomaney 'By the regions, for the regions' *The Guardian*, 27 July 2004.

¹¹¹ See Local Government Act 2000, Part 1.

¹¹² One of the implications of the introduction of elected Regional Assemblies would be to revise the two tier structure of local government. See M. Chisholm 'Reorganizing Two-Tier Local Government for Regional Assemblies' *Public Money and Management*, April 2004 pp.113-120.

¹¹³ Draft Bill on Regional Assemblies 2004: Policy Statement, para.23.

¹¹⁴ Regional Assemblies (Preparation) Act 2003 Part 2 s.13.

the financial markets.¹¹⁵ Second, additional funds could be raised by the Regional Assemblies through local authority council tax. This money would be collected on behalf of the regional assembly by local authorities in the region as part of existing arrangements for collecting council tax¹¹⁶ but this precepting was to be achieved without affecting non-domestic rate payers and thus imposing an extra burden on business. An elected assembly would have been allowed to set a higher charge to fund additional spending if the assembly considered such a charge desirable, but the controlling hand of Westminster was to remain very much in evidence. Central Government would have been able to limit the amounts through arrangements similar to the existing local authority council-tax capping regime.¹¹⁷ In fact, overall these proposals stand out as a modest divestment of power by central government next to other devolved systems of regional government. For example, in Italy¹¹⁸ following the amendment of the Italian Constitution all matters not specifically mentioned in the Constitution are now assigned to the ordinary Regions.¹¹⁹ In Italy central government no longer has control over the legitimacy of regional administrative decisions, and the potential for granting significant financial autonomy has been increased.

Approval by Referendum

A further criticism concerns the way a referendum requirement has been applied to these and other changes to the structure of regional government.¹²⁰ The implementation of the government's proposals for elected regional assemblies depended on a referendum being held in each of the designated regions to approve the principle.¹²¹ This requirement might appear to follow

¹¹⁵ *Your Region, Your Choice: Revitalising the English Regions*, 2002, Cm 5511 at 5.11 states that: 'In principle, we believe that elected assemblies should decide on their own levels of capital spending to be supported by borrowing and should be accountable to voters for their decisions, subject to the levels of borrowing being prudent'.

¹¹⁶ It has been estimated that the contribution of council tax-payers to the running costs of the assembly would have been equivalent to around five pence per week for a Band D council tax-payer in any region

¹¹⁷ See *Strong Local Leadership – Quality Public Services*: Part II, chap.4. For example, regional assemblies would have been subject to a prudential regime that is also intended for local government which includes reserve powers to set lower prudential limits in certain circumstances.

¹¹⁸ See, e.g. Law 142/1990 which introduced general legislation for local government; Law 81/1993 which allowed for the direct election of City Mayors and Provincial Presidents; Law 59/1997 which allowed the government to transfer administrative functions to the Regions, Provinces and Town Councils; Legislative Decree 56/2000 which set out the fiscal arrangements. P. Leyland, J. Frosini, C. Bologna 'Regional Government Reform: assessing the prospects for devolution' [2002] *Public Law* 242-252.

¹¹⁹ Art.117 of the Italian Constitution.

¹²⁰ As well as the substantive arguments regarding referendums mentioned here the North East referendum was also controversial because of experimentation with postal voting techniques which it was claimed were open to abuse.

¹²¹ See the Regional Assemblies (Preparations) Act 2003 which, for example, enables the Government to specify the region(s) able to hold a referendum on an elected regional assembly and to fix the date of the poll. It also allows the

the same pattern as adopted for devolution, since a referendum was necessary to secure approval for the respective models of devolution in the devolved parts of the UK. Although a test of public opinion has been built into the process to determine whether there is sufficient public support in favour of regional government, the use of referendums in the manner prescribed cannot easily be justified. First, there is no scope under the Act to hold a general referendum in England on the broad principle of democratically elected regional government. The decision to set up the referendum process on a region-by-region basis was because of the lack of political support for regional government in the most highly populated parts of South East England. Reading between the lines, the government wanted to avoid embarking on a process that could fail at the first hurdle, as would be the case if a national referendum were held to test opinion and that referendum went against the principle of regional government. Second, the Act allows the minister to approve the holding of referendums in some regions and not others. After the legislation was introduced John Prescott, the Deputy Prime Minister, announced that he intended that there would be three regional referendums in 2004¹²² and later decided to cancel the planned referendums in Yorkshire and the North West region. The rationale that has been offered for this incremental approach is that a limited advanced consultation process would indicate a reservoir of basic support. However, the fact that citizens in some parts of the UK should have the entitlement to be consulted on such an important issue, while others were not to have that right, simply because they live in a different part of the country, cannot be defended. Third, the introduction of regional government itself depends upon gaining approval by referendum on a region-by-region basis. If some regions were to vote in favour, and others voted against, an outcome which looked likely on the basis of views which were current when the proposals were conceived, there would be an asymmetrical distribution of powers which would be even more pronounced than under present arrangements. Such an imbalance is justified by the government on the grounds that the electorate in some regions might consider that the degree of political representation is already sufficient, while others chose more democratic accountability by opting for a regional assembly. However, this reasoning ignores the principal argument for introducing a layer of government for the English regions, namely, to redress the inequality of democratic representation and accountability that has been greatly exacerbated by devolution.¹²³ The regional government initiative fails to address this problem on a national basis in England.

Finally, referendums have become an increasingly important part of the constitutional consultation process, but there are no established rules or conventions to determine when they should be held. As one commentator has pointed out: 'The high-sounding nature of this principle [of seeking national support by referendum] is mitigated somewhat by the facts that that whether or not there is a referendum and what constitutes fundamental change will be

government to set out the questions to be asked and determine those eligible to vote in such a referendum.

¹²² These will be in the North East, Yorkshire and Humberside and the North West.

¹²³ R. Johnston & D. Rossiter 'Inequality of Representation in the United Kingdom: A Constitutional Mess?' 2002, *Political Quarterly* 158-171.

decided by the government'.¹²⁴ In the case of the English regions it has been argued that the minister under Regional Assemblies (Preparations) Act 2003 has been granted too much discretion in deciding whether to hold regional referendums. Nevertheless, the Act provided that in the event of there being a 'no' vote the minister is forbidden from calling another referendum in that region for seven years.¹²⁵ The overwhelmingly decisive failure of the referendum in the North East of England has not only confirmed the difficulty in promoting popular support for elected regional government, but it has sidelined the published proposals for the foreseeable future.

Conclusion: What Next for England?

The proposals for elected regional assemblies were criticised in the press and by opposition parties and campaign groups¹²⁶ and the adverse reception contributed to their decisive rejection in the referendum held in the North East region on 4 November 2004.¹²⁷ So what went wrong? Federal systems of government tend to be adopted as a constitutional solution in nations with a substantial territorial area and an internal diversity of traditions. On the other hand, devolution has been developed in the UK to satisfy the particular requirements of the home nations of Scotland, Wales and Northern Ireland and a strategic level of government has been restored for London (supported by a referendum). Systems of devolved government with elements in common with the UK experience have been introduced in other European nations such as Spain and Italy. By way of contrast, the English regions have been invented as territorial entities on the basis of the RDA's which were set up for economic and administrative reasons rather than as a response to local political pressure, or to the constitutional imbalance caused by devolution.¹²⁸ Although English citizens in some parts of England feel some regional identity, particularly in the North East and Yorkshire, dividing England into regions has been an exercise of drawing lines on a map: an approach more reminiscent of the scramble for empire in the nineteenth century. On its own such a geographical division is hardly a sound basis for establishing a popular system of devolved government and not one likely to restore flagging public confidence in politics at sub-national level at a time when turnout in local government elections throughout the UK has been declining spectacularly. Another danger is that the arrangements for local government, and now regional government, simply becomes a familiar partisan battleground. The seven Metropolitan Authorities which operated at a similar level to regional government, between central government and local

¹²⁴ B. Hadfield 'Devolution, Westminster and the English Question' [2005] *PL* 286-305 at p.295. In discussing how far reaching changes need to be in order to require referenda, Professor Hadfield points out that the Scottish Parliament Constituencies Act 2004, without holding a referendum, retained the number of MSPs.

¹²⁵ See Regional Assemblies (Preparations) Act 2003 s.6(2).

¹²⁶ See Website of North East 'No' Campaign. <http://www.northeastnocampaign.co.uk/about.html>.

¹²⁷ The planned referendums in the North West region and Yorkshire had already been postponed.

¹²⁸ J. Tomaney 'New Labour and the Evolution of Regionalism in England' in J. Tomaney and J. Mawson (eds.) *England: The State of the Regions*, Bristol, Policy Press, 2002.

government until 1985 were dispensed with for political reasons. The abolition was regarded as a cynical response by the [then Conservative] government. All the authorities had been under Labour control and many were pursuing public transport, economic development and general expenditure policies which were in conflict with the Westminster government headed by Mrs. Thatcher.¹²⁹

The elected Regional Assemblies would have been given few competences. They would have lacked fiscal autonomy and they would have been granted limited powers. Moreover, Whitehall would have retained a prominent continuing role after the introduction of regional assemblies. In other words, the proposals would not have amounted to devolution in an equivalent form to the arrangements elsewhere in the UK.¹³⁰ However, the mistake was not the proposal to create a relatively weak form of government based on the English regions, but the attempt to introduce a new system applying only in some regions. A scheme introduced unevenly would have increased the asymmetries of devolution rather than tackling the underlying constitutional and administrative issues.

If we revisit our earlier discussion and note the manner in which regional governance currently operates, it appears that there remains a very strong case for building on foundations that have already been laid in the form of RDAs and Regional Assemblies. The introduction of such bodies can be regarded as recognition by government that there is a 'multi-layered constitution' which involves cooperative and participatory techniques of governance. The players in the overlapping policy networks require coordination at the base of the pyramid in all the English regions. The challenge for the government, and eventually Parliament, in addressing this issue is to introduce a convincing model applicable to the entire English nation which allows the practice of governance to be adequately connected to democratic mechanisms of control. A new strata of elected government with responsibility for the strategic functions presently in the hands of RDAs and Regional Assemblies, supplemented with responsibility for additional policy areas with a significant strategic dimension is required. This is not the place to offer a detailed discussion of a specific set of proposals, but the form of government introduced for the London Mayor and Greater London Authority is the obvious model which should be adopted.¹³¹ After a national referendum to approve the principle on a nationwide basis, each region would have an elected full time Regional Mayor responsible for policy implementation. This elected office, probably supported by a small cabinet, would be held in check by a regional assembly elected by an additional member system. The Assembly would have powers to amend or approve strategies coming from the Regional Mayor and make budget amendments. The Regional Mayor, in common with the London Mayor, would have a high

¹²⁹ See M. Loughlin *Legality and Locality: The Role of Law in Central-Local Government Relations*, Oxford, Oxford University Press, 1996 at p.118. These authorities were introduced in the 1970's and were abolished in the mid 1980's.

¹³⁰ R. Cornes 'Devolution and England: What is on Offer' in N. Bamforth and P. Leyland *Public Law in a Multi-Layered Constitution*, Oxford, Hart, 2003 at p.130ff.

¹³¹ P. Hetherington 'Howard targets regional bodies' *The Guardian*, 11 November 2004.

profile and become the personification of the region following in the footsteps of Mayor Ken Livingstone in London.¹³² It is possible to envisage such a post emerging as an attractive regional position for politicians, either on the way up, or on the way down in their political careers. The Regional Mayor and cabinet, subject to the approval of the assembly, would be responsible for appointments to the RDA and other regional quangos. A national framework on this basis, although not equivalent to devolution, would achieve parity at regional level throughout England, and although the financial parameters would be variable according to the functions given to these bodies, such a system would be much less expensive to introduce than proposals for an elected Regional Assembly based on even the Welsh devolution model.¹³³ Of course, there would be many further questions to answer relating to such a system. For example, what precise changes to make regarding central and local government to allow the new regional level of government to be seamlessly integrated? Should non-elected stakeholders continue to be represented as part of an elected assembly and should non-elected stakeholders be granted voting powers? Should the regional assemblies have powers to make delegated legislation? What degree of financial autonomy should the new bodies be granted? How would the regional elected Mayor be coordinated with other local authorities?¹³⁴ In view of the adverse reception given to an elected assembly for the North East, there is no immediate chance of progress with such an agenda. However, looking a little further into the future, it is possible to envisage the enthusiastic support of influential political figures, such as an incoming Prime Minister, to promote a new scheme of elected government applying to all the English regions.

The present layer of regional governance, based on RDAs and Regional Assemblies, allows an indirect form of representation at regional level, as the Regional Assemblies consist of members nominated by local authorities in their region, but the central issue remains to be addressed, namely, improved accountability mechanisms. These are absent at this intermediate level of government. As an interim measure pending the introduction of elected regional assemblies, important quangos which have responsibility for regional development should be placed under democratic control in order to

¹³² It is acknowledged that there are arguments against going down the path of 'personality' politics. See P. Leyland 'The Modernisation of Local and Regional Government in England: Towards a New Democratic State' in J. Frosini, L. Pegoraro, T. Waller *Europe, Regions and local government in Italy, Spain and the United Kingdom*, Bologna, Libreria Bonomo editrice, 2003 at p.189ff.

¹³³ See G. Marquand and J. Tomaney, *Democratising England*, London Regional Policy Forum 2000 and C. Stevens 'English Regional Government' in *Devolution and British Politics*, Harlow, Longmans, 2004 who points out at p.253 that a strategic coordination model outlined in the White Paper would have a budget of £20 million, while a strategic-executive model would have a budget between £1.1-2.5 billion depending on the size of the region. The Welsh model on the other hand would cost between £6 - £14 billion depending on the size of the region,

¹³⁴ See M. Sandford and P. Hetherington, 'The Regions at the Crossroads: The Future of Sub-National Government in England' in A. Trench (ed) *The Dynamics of Devolution: The State of the Nations 2005*, Imprint Academic, Exeter, 2005 p.106.

reduce the potential for central government to use powers of patronage in making appointments without any democratic constraints. The Labour government has expressed a commitment to greater transparency in the way such appointments are made and favours equal representation of women and pro rata representation of ethnic minorities.¹³⁵ Until sufficient political backing exists for elected regional government, English regional quangos, in line with the position in Scotland, Wales, Northern Ireland and London, could be made directly accountable to their electorate. At local government elections as well as casting votes for councillors local electors should be given the opportunity to vote for candidates to serve on the RDA's and other important quangos operating at regional level.

¹³⁵ See *Quangos: Opening the Doors* (1998).

DISABILITY, DEGRADATION AND DIGNITY: THE ROLE OF ARTICLE 3 OF THE EUROPEAN CONVENTION ON HUMAN RIGHTS

*Anna Lawson, School of Law and Centre for Disability Studies,
University of Leeds**

Introduction

Traditionally, disability has not been regarded as a human rights issue. It has instead been viewed as a medical problem calling for medical solutions and for the support, through welfare or charity, of those who cannot be cured or corrected. The disability movement has insisted that disability is a human rights issue and that the attention of legislators and policy makers should be focussed on the physical, social, attitudinal, legal and other barriers which prevent disabled people participating in mainstream society.¹ In the language of the social model of disability,² it is societal barriers such as these which disable people who have physical, sensory, intellectual or other impairments.

The disability movement, then, calls for the dismantling of disabling societal obstacles and for the development of laws and policies which foster principles of inclusion and participation. Its goal is a society in which people with impairments are neither deliberately nor unthinkingly disabled by the

* This article originated in a paper ("Article 3 and Disabled People: Humanity, Degradation and Dignity") delivered at a conference ("Human Rights & Disability Discrimination: Exploring the value added by the ECHR and other sources of European Law") held in the University of Galway on 25 September 2004 the proceedings of which are due to be published jointly by the Disability Law & Policy Research Unit, Faculty of Law, NUI Galway and the Irish Equality Authority. I would like to record my thanks to Donncha O'Connell and the other organisers of that Conference, as well as to all its participants, for the opportunity to discuss some of the ideas considered here. I would also like to thank Caroline Gooding (Disability Rights Commission), Steven Wheatley (University of Leeds) and Colin Harvey (Queens University Belfast) for their very helpful comments on earlier drafts.

¹ See further, C. Barnes, *Disabled People in Britain and Discrimination: a Case for Anti-Discrimination Legislation* (London: Hurst/BCODP, 1991); C. Barnes and G. Mercer, *Disability* (Oxford: Polity Press, 2003); C. Barnes (ed.), *Implementing the Social Model of Disability: Theory and Research* (Leeds: The Disability Press, 2004); T. Degener, "Disabled Persons and Human Rights: The Legal Framework" in T. Degener and Y. Koster-Dreese (eds), *Human Rights and Disabled Persons: Essays and Relevant Human Rights Instruments* (Dordrecht/Boston/London: Martinus Nijhoff, 1995); J. Campbell and M. Oliver, *Disability Politics: Understanding our Past, Changing our Future* (New York: Routledge, 1996); M. Oliver, *The Politics of Disablement* (Basingstoke: Macmillan, 1990); M. Oliver, *Understanding Disability: From Theory to Practice*, (Basingstoke: Macmillan, 1996); M. Priestley, "Constructions and Creations: Idealism, Materialism and Disability Theory" (1998) 13 *Disability & Society* 75.

² See, e.g. Union of Physically Impaired Against Segregation, *Fundamental Principles of Disability* (London: UPIAS/Disability Alliance, 1976).

world around them: a society in which their individuality is valued and in which their dignity is respected. Human rights law has an essential role to play in the creation of such a society.

Although the protection of human dignity is a concern which runs throughout the European Convention on Human Rights 1950 (ECHR), it finds its clearest expression in Article 3, according to which:

“No one shall be subjected to torture or to inhuman or degrading treatment or punishment.”³

Potentially, then, Article 3 has a particularly important role to play in ensuring that people with impairments are treated as valued and respected members of society. It is my aim here to explore the scope of this potential role and the extent to which it has been developed to date.

While this discussion will be of relevance to disabled people in all countries in which the ECHR has effect, its prime focus will be the UK. The advent of the Human Rights Act 1998 has already generated a substantial body of UK Article 3 cases to add to those of the European Court of Human Rights (ECtHR).⁴ Although many of these UK cases have arisen in contexts far removed from disability, the principles developed in them have obvious relevance to disabled people.

It should be stressed that no precise definition underlies references to the term “disabled” in this article. One of the great strengths of the ECHR is that, unlike statutes such as the Disability Discrimination Act 1995,⁵ its protection is not conditional on compliance with any rigid definition of disability. The term will be used here simply to refer to people who have physical, mental, intellectual or sensory impairments (including those caused by serious illness).

In the next section I will provide a brief outline of the general scope of Article 3 and will attempt to draw out some of its aspects which have particular relevance for disabled people (despite the fact that these aspects have, to date, been developed largely in non-disability contexts). In the following section I will examine the way in which Article 3 has been used in cases (both of UK courts and of the ECtHR) concerning disabled people. That section will be divided into three parts reflecting three different types of living environment in which disabled people have found themselves and

³ Article 5 of the Universal Declaration of Human Rights 1948 is couched in very similar terms. Other relevant UN provisions are Articles 7 and 10 of the International Covenant on Civil and Political Rights 1966 and also the Convention against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment 1984. See generally, G. Quinn, “The Integrity of the Person: The Convention Against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment and Disability” in G. Quinn and T. Degener (eds), *Human Rights and Disability - the current use and future potential of United Nations Human rights instruments in the context of disability* (UN, November 2002), at www.ohchr.org/english/issues/disability/study.htm (last visited 3 August 2005).

⁴ For an analysis of the impact of these cases on various sectors of the population, see, C. Harvey (ed.), *Human Rights in the Community* (Oxford: Hart Publishing, 2005).

⁵ Section 1 of which defines a “disabled person”. See further, G. James, “The Meaning of Disability: Physical and Mental Impairment” [2002] ILJ 156.

which generate slightly different considerations for Article 3: first, prisons; second, other institutions and, finally, the home environment. The prison environment is addressed first because it is in that context that the Article 3 jurisprudence relating to disability has been most fully developed. I will argue that, although the cases demonstrate that Article 3 has begun to play a role in the promotion and protection of the dignity, safety and self-respect of disabled people in this country, its potential is currently under exploited, particularly in situations where the suffering is experienced by people living in their own homes. In the final substantive section I will consider the debate surrounding the issue of whether Article 3 does in fact offer protection to all or whether there are some disabled people who fall completely outside its reach.

Scope of Article 3

“Torture”, “Inhuman” and “Degrading”

The terms “torture”, “inhuman”, and “degrading” in Article 3 have not been precisely defined. Some general guidance, however, can be gleaned from cases such as *Ireland v UK*.⁶

“Torture” generally requires deliberate inhuman treatment, causing very serious and cruel suffering, which is inflicted for one of a number of purposes (including the extraction of information or confessions).⁷ “Inhuman” treatment or punishment generally involves intense physical or mental suffering.⁸ “Degrading” treatment or punishment, on the other hand, focuses on the notion of humiliation; being treatment which arouses in the victim “feelings of fear, anguish and inferiority” capable of humiliating and debasing them and of “breaking their physical and moral resistance”.⁹ For “inhuman” or “degrading” treatment, a deliberate intention to cause suffering or to humiliate and debase is an important, but not an essential, factor.¹⁰ In cases where there is no such intent, particular attention will be paid to the State’s response to the situation and its attempts to alleviate the relevant suffering.

Disabled people are, by no means, strangers to the deliberate infliction of cruelty or exploitation. This is recognized by Rule 9(4) of the UN Standard

⁶ Series A no 25 (18 Jan 1978); see also *The Greek Case* (1969) 12 YB 1 501.

⁷ See *Aydin v Turkey* (1997) 25 EHRR 251. See generally J. Cooper, *Cruelty: An Analysis of Article 3* (London: Sweet and Maxwell, 2003), 9-17; D. Feldman, *Civil Liberties and Human Rights in England and Wales* (Oxford: Oxford University Press, 2002), 245-246; and P.J. Duffy, “Article 3 of the European Convention on Human Rights” (1983) 32 ICLQ 316, 316-318.

⁸ See generally J. Cooper, *Cruelty: An Analysis of Article 3* (*ibid.*) 21-22; and P.J. Duffy, “Article 3 of the European Convention on Human Rights” (*ibid.*), 318.

⁹ *Pretty v UK* (2002) 35 EHRR 1, para.52. See generally J. Cooper, *Cruelty: An Analysis of Article 3* (*ibid.*), 22-23; P.J. Duffy, “Article 3 of the European Convention on Human Rights” (*ibid.*), 319-320; J. Vorhaus, “On Degradation: Degrading Treatment And Punishment” [2003] CLWR 65; and R.A. Duff, “Punishment, Dignity and Degradation” (2005) 25 OJLS 141, especially at 149-155.

¹⁰ *V v UK* [2000] 30 EHRR 121, para.71; *Peers v Greece* [2001] 33 EHRR 51, para.74.

Rules on the Equalisation of Opportunities for Disabled People 1993 which draws attention to the fact that they are “particularly vulnerable to abuse”.¹¹ Nevertheless, it seems probable that the suffering and humiliation which disabled people experience is more often the result of neglect or of failure to make adjustments in their favour or to provide them with necessary support. The Article 3 concepts of “inhuman” and “degrading” treatment are therefore likely to be of more assistance to them than is that of “torture”.

Article 3’s prohibition of torture and inhuman or degrading treatment or punishment is absolute, allowing no exceptions or limitations. The courts have consequently set a high threshold for entry into its realms. Suffering must reach a certain minimum level before it can amount to an Article 3 violation.¹²

As well as imposing a negative obligation on States to refrain from inflicting torture or inhuman or degrading treatment or punishment on its citizens, Article 3 has been held to require that they conduct appropriate investigations into possible breaches of Article 3 which come to their attention.¹³ It also imposes positive obligations which require States to take reasonable steps to protect those at a real risk of experiencing relevant suffering. These positive obligations widen the scope of Article 3 and considerably increase its potential usefulness to disabled people. They will therefore be considered in more depth below. First, however, some attention will be given to the concept of “treatment” used in Article 3.

“Treatment”

It should be stressed at the outset that, although many impairments may bring with them intense pain and suffering, Article 3’s negative prohibition against inflicting inhuman or degrading treatment will not be violated unless it can be shown that that suffering has, or may have been, exacerbated by treatment for which the authorities can be held responsible.¹⁴ Identification of such treatment is far from unproblematic. In *Pretty v UK*,¹⁵ Diane Pretty argued that her husband should be able to assist her to commit suicide, without fear of criminal prosecution, in order to spare her from the humiliation and degradation which she would otherwise suffer in the last stages of her terminal illness. She contended that the State’s refusal to refrain from prosecuting her husband, should he assist her, amounted to treatment which exacerbated her suffering and which therefore violated Article 3. This argument was rejected.¹⁶ The suffering in question was caused entirely by her medical condition and the refusal of the State to exempt her husband from

¹¹ See further, H. Brown, *Safeguarding Adults and Children with Disabilities Against Abuse* (Strasbourg: Council of Europe Publishing, 2002).

¹² See, e.g. *Pretty v UK* (2002) 35 EHRR 1, para.52. See generally J. Cooper, *Cruelty: An Analysis of Article 3* (n 7 above).

¹³ *Assenov v Bulgaria* (1998) 28 EHRR 652; see also A. Mowbray, “Duties of Investigation Under the European Convention on Human Rights” (2002) 51 ICLQ 435.

¹⁴ *Pretty v UK* (2002) 35 EHRR 1, para.52. See also A. Petain, “The Human Rights Dimensions of the Diane Pretty Case” [2003] CLJ 181.

¹⁵ *ibid.*

¹⁶ *ibid.*, at para.53.

criminal prosecution did not constitute “treatment” within the meaning of Article 3.

The ECtHR in *Pretty* distinguished the case before it from that of *D v UK*.¹⁷ There, the ECtHR had held that the UK would violate Article 3 if it implemented its decision to deport a prisoner dying of AIDS to St Kitts. This was because, if returned to St Kitts, there would have been no prospect of him receiving medical treatment or family support and there would consequently have been a real risk that he would die in extremely painful and distressing circumstances. “Treatment”, in *D*, could be located in the expulsion of the claimant from the country which would have exacerbated the suffering flowing from his naturally occurring illness.

In *Pretty* there is some reference to the fact that, in *D*, the expulsion would have constituted an “act” in contrast to the situation in *Pretty* itself where there was no such “act”.¹⁸ Nowhere, however, does the ECtHR stipulate that “treatment” must always take the form of an act rather than an omission.

In the UK, the statutory regime imposed on asylum seekers (which prevents them both from working¹⁹ and from receiving State support²⁰) has recently been held to constitute “treatment” within the meaning of Article 3.²¹ According to Phillips MR in *R v Sec of State for the Home Dept, ex p Q*:

“Treatment, . . . implies something more than passivity on the part of the State; but here, it seems to us, there is more than passivity. . . . The imposition by the legislature of a regime which prohibits asylum seekers from working and further prohibits the grant to them, when they are destitute, of support amounts to positive action directed against asylum seekers and not to mere inaction.”²²

Lord Scott, in *R v Sec of State for the Home Dept, ex p Adam, Limbuela and Tesema*,²³ was at pains to stress that there would not be any relevant “treatment” where destitution or homelessness could be attributed only to a failure of a State to adopt effective preventative measures. It was not the function of Article 3, in his view, “to prescribe a minimum standard of social support for those in need”²⁴ and consequently “treatment” required “something more than mere failure”.²⁵

¹⁷ (1997) 24 EHRR 423.

¹⁸ (2002) 35 EHRR 1, para.53.

¹⁹ Asylum and Immigration Act 1996, s.8.

²⁰ The Immigration Nationality and Asylum Act 2002, s.51(1).

²¹ *R v Sec of State for the Home Dept, ex p Q* [2004] QB 36 and *R v Sec of State for the Home Dept, ex p Adam, Limbuela and Tesema* [2005] UKHL 66. For criticism of this aspect of the decision in *Q*, see *Sec of State for the Home Dept v Limbuela and Others* [2004] QB 1440, paras.[115]-[117] *per* Carnwarth LJ.

²² [2004] QB 36, para.[56]; approved in *R v Sec of State for the Home Dept, ex p Adam, Limbuela and Tesema* [2005] UKHL 66, at para.[6] *per* Lord Bingham and para.[56] *per* Lord Hope.

²³ *ibid.*

²⁴ *ibid.*, at para.[66]. See also para.[7] *per* Lord Bingham.

²⁵ *ibid.*

It is unclear how this emerging test for “treatment” would operate in cases where suffering is exacerbated, not by a legislative scheme as such, but by a negligent failure to apply it. If, for instance, a disabled person’s suffering were exacerbated by the failure of their Local Authority to take any steps to provide them with support to which they were entitled, should the Authority’s failure to act (in breach of a statutory duty) be characterised as passivity or, alternatively, as some form of positive failing and therefore as relevant “treatment”? As will be demonstrated below, there is clear authority to the effect that Article 3 may be engaged where suffering has been caused or exacerbated by similar types of State omission or failure to act.²⁶ If the meaning of “treatment” is to be confined to positive acts, it is therefore important that “treatment” is not viewed as an essential pre-requisite of every Article 3 violation not involving torture or punishment. It may be argued, for instance, that it should not be required in cases where Article 3 is found to impose a positive obligation on States to take steps to prevent relevant suffering.²⁷

In any event, when faced with decisions as to whether or not there was any relevant “treatment” in a particular case, it is clear that judges should bear in mind the ECtHR’s observation in *Pretty* that the interpretation given to such terms must accord with the “fundamental objectives of the Convention and its coherence as a system of human rights protection”.²⁸ Consideration should therefore always be given to the consequences of a decision as to the existence of “treatment”. This, it is hoped, will reduce the risk of an overly rigid distinction between action and passivity becoming determinative of liability for human rights abuses.²⁹

Negative and Positive Obligations

Reference has already been made to the existence of both negative obligations to refrain from inflicting inhuman and degrading treatment and positive obligations to take measures to prevent Article 3 suffering occurring. In a commonly cited passage, the ECtHR in *Pretty* outlined this distinction as follows:

“50. An examination of the Court’s case-law indicates that Article 3 has been most commonly applied in contexts in which the risk to the individual of being subjected to any of the proscribed forms of treatment emanated from intentionally inflicted acts of State agents or public authorities It may be described in general terms as imposing a primarily negative obligation on States to refrain from inflicting serious harm on persons within their jurisdiction. However, in light of the fundamental importance of Article 3, the Court has reserved to itself sufficient flexibility to address the application of that Article in other situations that might arise . . .

²⁶ See the text accompanying nn.36 and 37 below.

²⁷ This appears to have been assumed to be the case in *R v Sec of State for the Home Dept, ex p Q* [2004] QB 36 at paras.[51]-[52].

²⁸ (2002) 35 EHRR 1, para.54.

²⁹ In *R v Sec of State for the Home Dept, ex p Adam, Limbuela and Tesema* [2005] UKHL 66 at para.[92] Lord Brown described this distinction as a “false dichotomy” reliance on which was “unhelpful”.

51. In particular, the Court has held that the obligation on the High Contracting Parties under Article 1 of the Convention to secure to everyone within their jurisdiction the rights and freedoms defined in the Convention, taken in conjunction with Article 3, requires States to take measures designed to ensure that individuals within their jurisdiction are not subjected to torture or inhuman and degrading treatment or punishment, including such treatment administered by private individuals”³⁰

Clear disability-related examples of treatment which might violate the negative prohibition of Article 3 may be envisaged. They would include the State authorisation of programmes under which people with mental impairments, who lack the ability to withhold consent, are subjected to medical experimentation; and the application to such people of unnecessarily harsh and invasive forms of restraint in institutional settings. Given the nature of the judicial guidance to date, however, any identification of disability-related examples of Article 3’s positive obligations must necessarily be more tentative.³¹

The ECtHR has declined to provide any definitive guidance as to the scope of positive obligations or as to their relationship with their more familiar negative counterparts.³² Carnwarth LJ has summarised the effect of the relevant ECtHR and UK case law as follows:

“The general proposition which emerges is unremarkable. It is that, in areas of state responsibility, where state action or inaction may lead to suffering sufficiently serious to engage Article 3, the relevant state agency has a duty to take reasonable steps to avoid it.”³³

In *Pretty v UK*³⁴ it was suggested that positive obligations invoked under Article 3 should “involve the removal or mitigation of harm by, for instance, preventing any ill-treatment by public bodies or private individuals or providing improved conditions or care”. Thus, in *A v UK*³⁵ a positive obligation was invoked to prevent ill-treatment at the hands of a private individual. The same objective underlies the obligation invoked in *Z v UK*,³⁶ though the provision of improved conditions or care was also relevant in that case.

Z, it is suggested, has particular significance for disabled people wishing to invoke a positive obligation requiring State authorities to implement existing policies effectively in order to ensure that their suffering does not cross the

³⁰ (2002) 35 EHRR 1.

³¹ See generally, K. Starmer “Positive Obligations Under the Convention” in J. Jowell and J. Cooper (eds), *Understanding Human Rights Principles* (Oxford: Hart Publishing, 2001); and A. Mowbray, *The Development of Positive Obligations Under the European Convention on Human Rights by the European Court of Human Rights* (Oxford: Hart Publishing, 2004).

³² *Plattform Artze fur das Leben v Austria* (1991) 13 EHRR 204.

³³ *Gezer v Sec of State for the Home Dept* [2004] EWCA (Civ) 1730, para.[46].

³⁴ (2002) 35 EHRR 1, para.55.

³⁵ 100/1997/884/1096 (23 September 1998).

³⁶ (2002) 34 EHRR 97.

Article 3 threshold. A Local Authority was there held to have violated Article 3 through its neglect or failure to act. The Council, which had been under a statutory duty to protect the four children concerned, had allowed them to remain in their home for four and a half years after discovering that they were at risk of serious neglect and also physical abuse. Thus, the Authority had violated Article 3, not by any deliberate act, but by a “failure of the system”.³⁷

The ECtHR has indicated that, while the negative prohibition of Article 3 is absolute and unqualified, its positive obligations are not.³⁸ The positive duty is to take only such steps as are reasonable to prevent ill treatment of which the authorities know or ought to have known. In the context of the analogous duty under Article 2, the ECtHR has explained that “such an obligation must be interpreted in a way which does not impose an impossible or disproportionate burden on the authorities” and that whether it is violated “is a question which can only be answered in the light of all the circumstances of any particular case”.³⁹ In the words of Lord Bingham, approved by the ECtHR in *Pretty*:

“It stands to reason that while States may be absolutely forbidden to inflict the prescribed treatment on individuals within their jurisdictions, the steps appropriate or necessary to discharge a positive obligation will be more judgmental, more prone to variation from State to State, more dependent on the opinions and beliefs of the people and less susceptible to any universal injunction.”⁴⁰

Unlike the negative obligation, therefore, a positive obligation would allow States an opportunity to advance a justification for their part in bringing about the suffering in question. This contrast between the absolute nature of the negative prohibition and the more qualified nature of the positive obligation creates an inevitable temptation to set up a clear bright-line dichotomy between them. In recent asylum seeker cases, however, the Court of Appeal has doubted the practical benefit of any such distinction to judges faced with decisions as to whether or not there has been an Article 3 violation in a particular case. While recognising that the categorisation of obligations as positive or negative constitutes “a valuable tool for analyzing the jurisprudence in this area”,⁴¹ it has described the distance between them as “in practice fairly narrow”,⁴² and the distinction as of limited utility,⁴³ as likely to give rise to “sterile arguments”,⁴⁴ and as “not establish[ing] the

³⁷ *ibid.*, at para.74.

³⁸ See *Pretty v UK* (2002) 35 EHRR 1, para.14, quoting from *R v DPP, ex p Pretty* [2002] 1 AC 800, para.[15] *per* Lord Bingham.

³⁹ *Osman v UK* 87/1997/871/1083 28 October 1998 at para.116.

⁴⁰ *R v DPP, ex p Pretty* [2002] 1 AC 800, para.[15].

⁴¹ *Gezer v Sec of State for the Home Dept* [2004] EWCA (Civ) 1730, para.[56] *per* Elias J.

⁴² *R v Sec of State for the Home Dept, ex p Q* [2004] QB 36, para.[52] *per* Phillips MR.

⁴³ *Gezer v Sec of State for the Home Dept* [2004] EWCA (Civ) 1730, para.[28] *per* Laws LJ.

⁴⁴ *ibid.*

touchstone by which the nature of the State's obligation can be defined".⁴⁵ These concerns have been echoed in the House of Lords, particularly by Lord Brown.⁴⁶ Indeed, the difficulty is well illustrated by the fact that, while Phillips MR⁴⁷ and Laws LJ⁴⁸ have confidently characterised the obligation at issue in *D* as a negative one (to refrain from inflicting inhuman and degrading treatment through deportation), Lord Brown⁴⁹ has insisted that it is "essential" that such an obligation should be acknowledged to be a positive one (to continue the provision of medical treatment in the UK to the would-be immigrant).

Attempts have been made to articulate the distinction between cases in which justifications may and may not be advanced in terms other than those of positive and negative obligations. In *R v Sec of State for the Home Dept, ex p Adam, Limbuela and Tesema*,⁵⁰ for instance, Lord Hope suggested that the key distinction was between cases in which the State was directly responsible for the relevant suffering and ones in which it was only indirectly responsible.⁵¹ Similarly, Baroness Hale suggested that it was between cases in which the suffering had been inflicted by the State itself and ones in which the State should have intervened in order to protect a person from suffering inflicted by others.⁵² Laws LJ has offered an alternative categorisation based on the distinction between cases of deliberate State violence and others in which State acts or omissions have exposed a person to relevant suffering. Despite the centrality of this distinction to Laws LJ's analysis, he has depicted the range of Article 3 obligations as a spectrum.

Lord Justice Laws' Article 3 spectrum

In *Sec of State for the Home Dept v Limbuela and others*⁵³ and *Gezer v Sec of State for the Home Dept*,⁵⁴ Laws LJ set out his spectrum analysis of Article 3 obligations. At its heart, as mentioned above, lies a distinction between two types of case which, according to his Lordship is, "well recognised in the jurisprudence, but whose importance has perhaps not always been fully unraveled".⁵⁵ In category (a) lies the deliberate use of unlawful violence by State officials.⁵⁶ Category (b), on the other hand,

⁴⁵ *ibid.*, at para.[56] *per* Elias J.

⁴⁶ In *N (FC) v Sec of State for the Home Dept* [2005] 2 AC 296, para.[88] and *R v Sec of State for the Home Dept, ex p Adam, Limbuela and Tesema* [2005] UKHL 66, paras.[91]-[95].

⁴⁷ *R v Sec of State for the Home Dept, ex p Q* [2004] QB 36, para.[54].

⁴⁸ *Sec of State for the Home Dept v Limbuela and Others* [2004] QB 1440, para.[64].

⁴⁹ *N (FC) v Sec of State for the Home Dept* [2005] 2 AC 296, para.[88].

⁵⁰ [2005] UKHL 66.

⁵¹ *ibid.*, paras.[48] and [55].

⁵² *ibid.*, para.[77].

⁵³ [2004] QB 1440, paras.[69]-[70].

⁵⁴ [2004] EWCA (Civ) 1730.

⁵⁵ *Sec of State for the Home Dept v Limbuela and others* [2004] QB 1440, para.[59].

⁵⁶ In *Gezer v Sec of State for the Home Dept* [2004] EWCA (Civ) 1730, para.[24], he added ambiguously that this category might also include "other grave ill-treatment". However, there is no elaboration of this and the judgements do contain a clear emphasis on the significance of violence.

consists of “acts or omissions by the State which expose the claimant to suffering inflicted by third parties or by circumstance”.⁵⁷

Both categories contain subdivisions. Category (a) contains, first, authorised State violence and, second, unauthorised violence by State officials. Such unauthorised violence may take the form either of the use of excessive force in the discharge of authorised State business or of violence which has “no colour” of authority at all.⁵⁸ Category (b) breaks down into cases in which the suffering is the result of violence inflicted by non-State officials, on the one hand and where it is the result of circumstances resulting from State acts or omissions, on the other. These different scenarios, according to Laws LJ, represent different points on a spectrum.

The paradigm case of Article 3 violation is that of State authorised violence which, subject to a limited number of very clearly defined exceptions (relating, for instance, to the use of reasonable force in the execution of arrests) is totally prohibited by the Article and cannot be justified. In these cases Article 3’s prohibition constitutes “the law’s first and only word”.⁵⁹ At the other end of the spectrum lie cases such as *D v UK*⁶⁰ in which the suffering has been caused, not by violence, but by circumstances created by acts or omissions of the State in pursuance of lawful policy. Article 3 may, on occasion, be violated in these cases but only if the suffering is exceptional and if allowing the claim would hold a fair balance between the interests of the particular claimant and those of the general public.⁶¹ In such cases, Article 3 is “no more nor less than the law’s last word”, operating as “a safety net, confining the State’s freedom of action only in exceptional or extreme cases”.⁶² Between these two ends of the spectrum, in relation to the State’s obligation to prevent, investigate and punish unauthorised violence for instance, the law allows States “a measure of judgment in the elaboration of measures to serve the aims”.⁶³ This measure of judgment, however, would be much narrower than it would be in cases at the *D v UK* end of the spectrum.

According to Laws LJ, although there will be some point which divides State acts or omissions violating Article 3 from those which do not, there is no “brightline rule” for determining exactly where this point lies.⁶⁴ In his words:

⁵⁷ *Sec of State for the Home Dept v Limbuela and others* [2004] QB 1440, para.[59].

⁵⁸ *ibid.*, at para.[65].

⁵⁹ *ibid.*, at para.[77].

⁶⁰ (1997) 24 EHRR 423.

⁶¹ Thus, according to the House of Lords in *N v Sec of State for the Home Dept* [2005] 2 AC 296, *D* has been narrowly confined by subsequent ECtHR cases to situations in which the person who would otherwise be deported is actually dying and the potential receiving country offers no relevant medical treatment at all. Expansion beyond these narrow confines would impose too heavy a burden on citizens of the expelling country, the immigration policies of which might be seriously undermined. For discussion of the Court of Appeal’s decision in *N*, see C. Sawyer, “Insufficiently Inhuman: Removing AIDS Patients from the UK” [2004] JSWFL 281.

⁶² *Sec of State for the Home Dept v Limbuela and Others* [2004] QB 1440, para.[77].

⁶³ *ibid.*

⁶⁴ *ibid.*, at para.[71].

“The point is at the place between cases where government action is justified notwithstanding the individual’s suffering, and cases where it is not. Various factors will determine where this place is to be found. They will include the severity of the threatened suffering, its origin in violence or otherwise, and the nature of the government’s reasons or purpose in acting as it does.”⁶⁵

Laws LJ stressed that the division between categories (a) and (b) should not be equated with the distinction between negative and positive obligations.⁶⁶ Many scenarios, in his view, would give rise to both negative and positive obligations.

“The State’s duty to protect individuals from Article 3 ill-treatment will, depending on the circumstances, sometimes involve refraining from action, and sometimes involve taking action. In the context of the duty to protect the difference is serendipitous. But in category (a) cases – and especially their most heinous instance, State-sponsored violence – it is misleading to speak merely of a duty to protect: such things are, simply, absolutely forbidden.”⁶⁷

The current status of Laws LJ’s spectrum analysis is somewhat uncertain. It was approved by the other members of the Court of Appeal in both *Limbuela* and *Gezer* and by the House of Lords in *N*.⁶⁸ In the more recent case of *R v Sec of State for the Home Dept, ex p Adam, Limbuela and Tesema*,⁶⁹ however, Lord Hope and Baroness Hale expressed disquiet about it and concern that it lacked a solid grounding in the jurisprudence of the ECtHR.⁷⁰ They stressed that issues of proportionality or justification had no role at all to play in cases where the State was directly responsible for the relevant suffering, even if that was as a result of what Laws LJ had termed “legitimate government policy” not involving violence.⁷¹ Lord Brown, on the other hand, expressed broad agreement with Laws LJ’s approach because it highlighted “the many different considerations in play and the need in all but the clearest cases ‘to look at the problem in the round’”.⁷² Thus, had the destitution of asylum seekers not been intended by the government, Lord Brown would have regarded issues of proportionality as relevant to the assessment of whether Article 3 had been violated albeit that, on the facts, he would have regarded the hardship as outweighing the particular policy objectives.⁷³ He considered it unnecessary and unhelpful, in any Article 3 case, to have to categorise the State’s obligation as negative or positive in

⁶⁵ *ibid.*

⁶⁶ *Gezer v Sec of State for the Home Dept* [2004] EWCA (Civ) 1730, para.[27].

⁶⁷ *ibid.*, at para.[29].

⁶⁸ [2005] 2 AC 296, para.[87].

⁶⁹ [2005] UKHL 66.

⁷⁰ *ibid.*, paras.[53] and [77]-[78] respectively.

⁷¹ See, in particular, *ibid.*, para.[55] *per* Lord Hope.

⁷² *ibid.*, para.[89].

⁷³ *ibid.*, para.[101].

order to determine whether it would be permissible to consider issues of proportionality or justification.⁷⁴

Disability-Specific Article 3 Developments in Different Living Environments

Prisons

As already mentioned, in assessing whether the minimum level of severity has been attained for treatment or punishment to qualify as “degrading”, the court must take into account all the circumstances of the case, including the health, age and sex of the victim.⁷⁵ This stress on the circumstances of the particular victim has been used by judges to require States to pay due regard to any impairments (physical or mental) which a prisoner might have in deciding whether Article 3 has been violated.

In *Price v UK*,⁷⁶ which is perhaps the most important Article 3 case concerning disabled people, the UK was found to have inflicted degrading treatment on Ms Price by its failure to accommodate the particular needs arising from her impairment. As a result of thalidomide, all four of her limbs were fore-shortened and she had a number of associated health difficulties, including kidney problems. Her refusal to answer questions in County Court debt recovery proceedings resulted in her spending three nights in prison. She was refused permission to take the battery charger for her wheelchair to prison; she had to spend the first night in a cell which was dangerously cold for her and contained a bed she was unable to use; she had to be assisted in using the toilet by male staff (having been left sitting on the toilet for three hours on one occasion until she gave up hope of being assisted by a woman); and, at the end of her sentence, she required catheterisation due to lack of fluid and to urine retention caused by difficulties in using the toilet facilities.

The ECtHR ruled that Ms Price’s Article 3 rights had been infringed, observing that:

“There is no evidence in this case of any positive intention to humiliate or debase the applicant. However, the Court considers that to detain a severely disabled person in conditions where she is dangerously cold, risks developing sores because her bed is too hard or unreachable, and is unable to go to the toilet or keep clean without the greatest of difficulty, constitutes degrading treatment.”⁷⁷

Judge Greve, who delivered a separate concurring opinion, added that:

⁷⁴ *ibid.*, paras.[90]-[92].

⁷⁵ *Ireland v UK* Series A no. 25 (18 January 1978), para.162.

⁷⁶ (2001) 34 EHRR 1285. For a similar decision in the UN context, see, *Hamilton v Jamaica*, communication No 616/ 1995, views adopted by the Human Rights Committee on 28 July 1999 (CCPR/C/66/D/616/1995); discussed in T. Degener, “Disability and Freedom: The International Covenant on Civil and Political Rights” in G. Quinn and T. Degener (eds), *Human Rights and Disability – the current use and future potential of United Nations human rights instruments in the context of disability* (n.3 above), para.4.1.4.

⁷⁷ *ibid.*, at para.30.

“In a civilised country like the United Kingdom, society considers it not only appropriate but a basic humane concern to try to ameliorate and compensate for the disabilities faced by a person in the applicant’s situation. . . . The applicant’s disabilities are not hidden or easily overlooked. It requires no special qualification, only a minimum of ordinary human empathy, to appreciate her situation and to understand that to avoid unnecessary hardship - that is, hardship not implicit in the imprisonment of an able-bodied person - she has to be treated differently from other people because her situation is significantly different.”⁷⁸

According to other Article 3 cases, the illness or impairment of a prisoner may, on occasion, require that they be removed from the mainstream prison environment, either by being released or by being transferred into other conditions (such as those of a hospital). Thus, in *Mouisel v France*,⁷⁹ the ECtHR held that Article 3 had been contravened by the continued detention of a prisoner suffering from leukaemia and undergoing chemotherapy. On his frequent trips to the hospital for treatment he was chained and handcuffed. He alleged that during the chemotherapy sessions his feet had been chained and one of his wrists handcuffed to the bed. He also alleged that, following his attempt to reduce the pain of the treatment by altering the speed of the drip, prison guards physically restrained him and dragged him through the hospital corridors by his chains.

Similarly, in *Napier v Scottish Ministers*,⁸⁰ Lord Bonomy ruled that the detention of a prisoner, who had eczema, in the conditions which then existed in the mainstream prison environment constituted degrading treatment. Robert Napier had been required to spend between twenty and twenty-three hours a day in his cell; to share that cell with a series of cell-mates even though it was designed for only one; and, most significantly, had been granted extremely limited access to washing and toilet facilities. The “slopping out” regime to which Napier was subjected meant using a bucket or bottle in the cell instead of a toilet, emptying these receptacles alongside other prisoners in chaotic conditions, and having extremely limited opportunities to wash: so limited that it was impossible for Napier to take adequate care of his skin which consequently became infected, ulcerated and flaky. Interestingly, though Lord Bonomy regarded the existence of Napier’s eczema as “crucial”⁸¹ to the success of his Article 3 claim, he never described him as “disabled” and made no reference to the *Price* decision. This illustrates the ability of Article 3 to respond to the particular circumstances of disabled people without imposing on them any threshold requirement in the form of a specific definition of “disability”. It thus stands in refreshing contrast to the Disability Discrimination Act 1995 where failure to satisfy the

⁷⁸ *ibid.*

⁷⁹ Application No. 67263/01 (14 November 2002).

⁸⁰ [2004] SLT 555. See, for further discussion of this case, A. Lawson and A. Mukherjee, “Slopping Out in Scotland: The Limits of Degradation and Respect” [2004] EHRLR 645.

⁸¹ *ibid.*, at para.[76].

definition of disability has caused more cases to flounder than has any other reason.⁸²

As well as demonstrating the need, on occasion, to modify regimes or conditions for disabled prisoners, these cases illustrate the need to provide them with adequate medical care. This requirement is also, and more directly, demonstrated by *Keenan v UK*.⁸³ Keenan was known to have schizophrenia and to be a suicide risk. Following a change in medication, he assaulted two prison guards and, without proper psychological assessments of the impact of the move on his health, was sent from the hospital block to the punishment block for a period of solitary confinement. During this time, Keenan committed suicide. The lack of adequate medical records covering the period in question indicated that, during it, his condition had not been adequately monitored. Similarly, in *McGlinchey v UK*,⁸⁴ it was found that Article 3 had been breached by a failure to provide appropriate care to a prisoner who had severe asthma and was dying from heroin addiction.

The need to provide adequate medical care to prisoners also underlies the decision in *D v UK*⁸⁵ which, as mentioned above, concerned the deportation of a prisoner dying of AIDS from the UK to St Kitts where he would have had no prospect of receiving medical care or family support. However, the fundamental issue in *D* and other such cases (which are not confined to the prison context) is the extent to which Article 3 operates to prevent States applying their general immigration policies to deport sick or disabled people to countries in which they are unlikely to receive adequate care.⁸⁶ A detailed examination of such cases would therefore stray beyond the scope of this article, the prime focus of which is the relevance of Article 3 to the lives of disabled people who are entitled to remain living in the UK.

In the prison context, then, Article 3 cases have played an important role in establishing minimum acceptable standards for the treatment of disabled people. They have drawn attention to the additional suffering which a regime may cause to a disabled person and stressed the importance of making appropriate adjustments in order to take account of their particular

⁸² S. Leverton, *Monitoring the Disability Discrimination Act 1995 (Phase 2)* (London: Stationary Office, 2002), 14.

⁸³ (2001) 33 EHRR 38.

⁸⁴ Application No. 00050390/99 (29 April 2003). See also, *R v Sec of State for the Home Dept, ex p Wright and Bennett* [2001] EWHC (Admin) 520.

⁸⁵ (1997) 24 EHRR 423.

⁸⁶ The issue was considered in some depth by the House of Lords in the non-prison case of *N (FC) v Sec of State for the Home Dept* [2005] 2 AC 296. According to it, the deportation of a would-be immigrant, who had a serious illness for which they were receiving medical care in the UK, was likely to ground an Article 3 violation only where that person was actually in the terminal stages of the disease at the time of the application and only then if the country to which they would otherwise be deported offered no relevant medical care at all. Thus, in *N*, Article 3 was not held to be violated by the deportation to Uganda of a woman who had advanced AIDS, but whose condition had been stabilised by treatment in the UK, where medical treatment was technically available in Uganda (despite the fact that it was extremely unlikely that the claimant would be able to access that treatment in practice and that, without it, her life expectancy would be reduced from decades to a year or two).

circumstances. An undoubtedly significant factor in these cases is the relative powerlessness of prisoners over their environment and the high degree of control which the State exercises over all aspects of their lives. Although this factor is also present (to some degree) where disabled people are required to live in residential institutions in order to receive adequate care, its absence from scenarios in which disabled people live at home seems (as will be considered below) to have contributed to an unwillingness to extend Article 3 protection to such cases.

Residential Institutions

Inadequate Care

Many disabled people continue to live in residential institutions, such as psychiatric hospitals, homes for the elderly and other long-stay institutions. Indeed, although government policy is to encourage independent living in the community, there is some indication that the number of disabled people living in institutional care increased between 1997 and 2002.⁸⁷ While some of these institutions undoubtedly provide exemplary regimes, many do not.⁸⁸ Jenny Watson's 2002 report⁸⁹ documents disturbing conditions in a number of such institutions. It includes accounts of practices and policies which appear to reflect a complete absence of concern for the dignity and well-being of residents. The following is an example:

“An agency worker told us about going into a home at breakfast time. She was instructed to get the residents up and onto their commode. She was then told to feed them breakfast. When she started to get the residents off their commodes first she was stopped. The routine of the home was that residents ate their breakfast while sitting on the commode and the

⁸⁷ Cabinet Office Strategy Unit *Improving the Life Chances of Disabled People* (London: Stationary Office, 2005), para.4.4.

⁸⁸ For a shocking example drawn from elsewhere in Europe, see Amnesty International's 2002 report on Bulgaria. It contains the following account: “In the Sanadinovo home for mentally disabled women, those who ‘misbehaved’ were held in a cage made of two brick walls with iron bars and wire on the remaining two sides. In October, an AI delegate witnessed six women being held in this 3m x 1.5m space. They looked withdrawn and vacant and appeared non-aggressive. The cage was full of urine and faeces and the women were covered in filth. One woman was naked from the waist down and many sores were visible on her skin. It was not possible to establish how long anyone would be subjected to this form of seclusion as no records of this appeared to be maintained.” See “Amnesty International: Europe – Bulgaria 2002” at <http://web.amnesty.org/web/ar2002.nsf/eur/bulgaria> (last visited 3 August 2005). See also Amnesty International press release “Bulgaria: Disabled women condemned to ‘slow death’”, AI-index: EUR 15/002/2001; and, more generally, O. Lewis, “Mental Disability Law in Central and Eastern Europe: Paper, Practice, Promise” (2002) 8 *Journal of Mental Health Law* 293.

⁸⁹ *Something for Everyone: The impact of the Human Rights Act and the Need for a Human Rights Commission* (London: British Institute of Human Rights, 2002). See also, H. Brown, *Safeguarding Adults and Children with Disabilities Against Abuse* (n.11 above), especially paras.2.1.3 and 3.5.

ordinary men and women who worked there had come to accept this as normal.”⁹⁰

The general tenor of the accounts documented in this report is captured by the following extract:

“In relation to care homes, people have their time organised for them. The sort of things you want to do in your own home can’t happen. Things like transport provision are usually for the convenience of the provider, not the user. We know of cases [in residential settings] where people have been left in their own excrement, where they are put to bed at five o’clock in the evening, where they are not allowed to watch TV in the evenings.”⁹¹

Rowena Daw, writing in 2000, provides another account of the type of suffering and degradation which may be experienced by a disabled person living in residential care:

“A deaf woman who signed was a long-term resident in a residential home. She was not given communication support because her notes explaining her need were lost. Lacking the means of communication she developed behavioural difficulties and was then misdiagnosed as of low intelligence. She was kept in total isolation for 30 years until the home was closed down and the truth was discovered. She had by then lost her ability to sign.”⁹²

Living conditions in such institutions, then, have resulted in intense suffering and humiliation for a considerable number of disabled people. In the words of Gerard Quinn, writing in the international context:

“It is undeniable that torture, and the less severe but equally abominable phenomena of cruel, inhuman or degrading treatment or punishment, have lurked quietly behind the closed doors of mental and physical disability institutions.”⁹³

The requirements of Article 3, it is suggested, should not be applied any less strictly in this context than in that of prisons. Reference is often made to the peculiar vulnerability of prisoners to abuse, given the control imposed on all aspects of their lives.⁹⁴ This vulnerability also characterises disabled residents living in institutions providing them with care on which they are

⁹⁰ *ibid.*, at 49. See also A. Clarke and L. Bright, *Showing Restraint: Challenging the Use of Restraint in Care Homes* (London: Counsel & Care, January 2002).

⁹¹ J. Watson, *Something for Everyone: The Impact of the Human Rights Act and the Need for a Human Rights Commission* (n 89 above), 37 (quoting from the submission of Andy Rickell, British Council of Disabled People).

⁹² *The Impact of the Human Rights Act on Disabled People* (London: Disability Rights Commission and Royal National Institute for Deaf People, 2000), 52.

⁹³ G. Quinn, “The Integrity of the Person: The Convention Against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment and Disability” (n 3 above), para.6.1.

⁹⁴ See, e.g. *Anguelova v Bulgaria* Application No 38361/97 (13 June 2002), para.110; and S Livingstone, T. Owen and A. MacDonald, *Prison Law* (Oxford: Oxford University Press, 2003) para.2.01.

totally dependent. Indeed, in psychiatric hospitals, many of them (like prisoners) will be there expressly against their will, and in other institutions, many will be there only because there is no alternative. This vulnerability was recognised by the ECtHR in *Herczegfalvy v Austria*⁹⁵:

“The Court considers that the position of inferiority and powerlessness which is typical of patients confined in psychiatric hospitals calls for increased vigilance in reviewing whether the Convention has been complied with.”

Thus it seems that residents of these institutions would, in theory, be able to use Article 3 to call for at least the same minimal level of care as that required in the prison cases. Indeed, it has been suggested by the Court of Appeal⁹⁶ that inadequate or inappropriate conditions of detention might be more likely to amount to inhuman or degrading treatment in the case of a patient than in that of a prisoner. This is because the detention of a prisoner is an end in itself whereas that of a patient is a means to an end: the end being the provision of care and, in some cases, the restoration of health. Article 3, it was suggested, may be more likely to be engaged where the conditions in question undermine the very purpose of the detention.

Nevertheless, the suffering of disabled people in residential institutions has given rise to surprisingly few Article 3 cases.⁹⁷ Possible explanations for this emerge clearly from even the most cursory glance through Jenny Watson’s report. Residents who do complain about inadequate care are sometimes faced with expulsion from the home: an extremely daunting prospect for those with no real alternative accommodation and one likely to deter the bringing of complaints.⁹⁸ Fear of consequences other than expulsion, such as victimisation by care providers, may also prove influential.⁹⁹ Even where complaints are made, the evidence of disabled people (particularly those with learning difficulties or mental illness) is often regarded as insufficiently cogent to found disciplinary actions or criminal proceedings.¹⁰⁰

⁹⁵ (1992) 15 EHRR 437, paras.82 and 110.

⁹⁶ *R v Mersey Care NHS Trust, ex p Munjaz, R v Airedale NHS Trust, ex p S* [2003] 3 WLR 1505, para.[56].

⁹⁷ See generally, L. Clements and J. Read, “The Dog that Didn’t Bark: The Issue of Access to Rights under the European Convention on Human Rights by Disabled People” in A. Lawson and C. Gooding (eds.), *Disability Rights in Europe: From Theory to Practice* (Oxford: Hart Publishing, 2005); and G. Quinn, “The Integrity of the Person: The Convention Against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment and Disability” (n.3 above), para.61.

⁹⁸ This is illustrated by the following extract (*Something for Everyone: The Impact of the Human Rights Act and the need for a Human Rights Commission* (n.89 above), 48): “[X] has been asked to leave her residential care home because she complained about a member of staff who would not attend to her in the night. She wanted to go to the toilet so she rang the buzzer but the staff member didn’t come for about half an hour, by which time it was too late. The care worker left her in her wet night clothes, took the buzzer from her and threw it across the room”.

⁹⁹ K. Simons, *I’m Not Complaining, but . . .* (York: Joseph Rowntree Foundation, 1995).

¹⁰⁰ This is illustrated by the following extract (*Something for Everyone: The Impact of the Human Rights Act and the Need for a Human Rights Commission* (n.89 above), 41, quoting from the submission of Rachel Hurst, Disability Awareness in Action): “We’ve got a case of a young black man, who’s severely autistic,

Unwanted Medical Treatment

An issue which has come before the ECtHR and UK courts on a number of occasions is the question of whether medical treatment or force-feeding, provided against the will of a person confined in an institution and lacking the capacity to consent, constitutes inhuman or degrading treatment. The leading case is *Herczegfalvy v Austria*.¹⁰¹

Mr Herczegfalvy, who was a psychiatric patient and lacked full mental capacity according to Austrian law, was isolated from other residents and attached with handcuffs and other straps to a security bed for a fortnight. Despite his struggles and protests (which resulted in him suffering the loss of teeth, broken ribs and bruising)¹⁰², food and neuroleptic medication were administered. He argued that this treatment had contributed to the worsening of his condition and constituted inhuman and degrading treatment contrary to Article 3. The Austrian government argued that the treatment in question had been necessary in order to preserve his health, pointing out that:

“[It] was only his resistance to all treatment, his extreme aggressiveness and the threats and acts of violence on his part against the hospital staff which explained why the staff had used coercive measures including the intramuscular injection of sedatives and the use of handcuffs and the security bed. These measures had been agreed to by Mr. Herczegfalvy’s curator [guardian], their sole aim had always been therapeutic, and they had been terminated as soon as the state of the patient permitted this.”¹⁰³

The Court, in ruling that there had been no breach of Article 3, laid down the “general rule” that “a measure which is a therapeutic necessity cannot be regarded as inhuman or degrading” and that, consequently, established principles of medicine would be “in principle decisive in such cases”.¹⁰⁴ It added, however, that: “The Court must nevertheless satisfy itself that the medical necessity has been convincingly shown to exist”.¹⁰⁵ In *R v M and Others, ex parte N*,¹⁰⁶ the Court of Appeal stressed that the “convincingly shown” requirement imposed a high standard of proof on the State but that it should not be equated with the criminal standard of “beyond reasonable doubt”.

we’ve got medical evidence that he was beaten black and blue in the home, his grandmother took him away. The CPS said initially they would take the case on, but then they withdrew, because they said he couldn’t give evidence”.

¹⁰¹ (1992) 15 EHRR 437. See also *X v Federal Republic of Germany* (1984) 7 EHRR 152 where the force-feeding of a prisoner on hunger strike was held not to violate Article 3; and *R v Responsible Medical Officer, Dr G and . . . Dr W, ex p PS* [2003] EWHC (Admin) 2335 where the provision of medical treatment to a competent patient against their will was held not to contravene Article 8.

¹⁰² Para.31.

¹⁰³ Para.81.

¹⁰⁴ Para.82.

¹⁰⁵ *ibid.*

¹⁰⁶ [2003] 1 WLR 562, paras.[17]-[18].

In *R v Dr S and Others, ex parte B*,¹⁰⁷ Charles J considered the application of *Herczegfalvy* to UK law in some depth. In so doing he reached two conclusions, both of which facilitate the application of unwanted medical treatment to people detained under the *Mental Health Act 1983* (MHA). First, he held that *Herczegfalvy* did not require a therapeutic necessity to be demonstrated in all cases of unwanted medical treatment. Such a necessity would be required only where the proposed treatment was likely to cause the patient suffering so severe as to engage Article 3. In decisions as to which side of this line a case might fall, according to Charles J, the evidence of the doctor caring for the patient (and presumably wishing to apply the treatment) would be of particular significance.¹⁰⁸ Second, he held that although the ECtHR in *Herczegfalvy* had been concerned with cases in which patients were “entirely incapable of deciding for themselves”,¹⁰⁹ the principles it laid down were also applicable to cases in which patients were capable of deciding for themselves but to whom Part IV of the *Mental Health Act 1983* authorised the administration of treatment contrary to their wishes.¹¹⁰

The therapeutic necessity test itself is not unproblematic, as has been convincingly argued by Oliver Lewis.¹¹¹ He describes it as “highly deferential” and draws attention to the difficulty which people deemed to lack mental capacity are likely to face in disproving medical necessity. Persuading judges to trust their testimony where it contradicts that of professional doctors is not likely to be easy. This problem is exacerbated by the decision in *B* that in some cases of unwanted medical treatment, to be identified largely by the responsible doctor, there will be no need to show that there was any therapeutic necessity for the intervention. It may also be intensified by the fact that, in the ECtHR, a violation of Article 3 must generally be established beyond reasonable doubt,¹¹² (although there is now some UK authority to the effect that the relevant standard is the usual civil standard of a balance of probabilities).¹¹³ Against this background, as Lewis notes, it is disturbing to learn that in recent years, in a number of European countries (though admittedly not the UK), doctors have authorised practices such as the electroconvulsive therapy of mentally-ill people without anaesthesia or muscle relaxants¹¹⁴ and their confinement in caged-beds (sometimes for months or years).¹¹⁵ It would not be difficult to challenge the

¹⁰⁷ [2005] EWHC 1936.

¹⁰⁸ *ibid.*, paras.[60]-[70].

¹⁰⁹ (1992) 15 EHRR 437, para.82.

¹¹⁰ [2005] EWHC 1936, para.[58]. See also *Re B* [2005] EWHC 86, paras.[109]-[112] *per* Silber J.

¹¹¹ “Protecting the Rights of People with Mental Disabilities: The European Convention on Human Rights” [2002] 9 *European Journal of Health Law* 293.

¹¹² See generally *Ireland v UK* (1978) 2 EHRR 25 paras.160-161.

¹¹³ *Napier v Scottish Ministers* [2005] UKHRR 268.

¹¹⁴ See, e.g. *Bulgarian Helsinki Committee Inpatient psychiatric care in Bulgaria and human rights* (Sofia: Bulgarian Helsinki Committee, 2001) reporting that this practice was carried out in at least eight psychiatric institutions in Bulgaria and that it is prescribed by twenty per cent of Bulgarian psychiatrists.

¹¹⁵ See G. Gombos, E. Kismrdi, K. Petr, *The Human Rights of Patients in Social Care Homes for the Mentally Ill* (Budapest, Hungarian Mental Health Interest Forum, September 2001) reporting that caged beds were used in eight of the fifty-two Hungarian institutions studied; see also “Disabled People Padlocked in

therapeutic necessity of these particular practices as they have been condemned as inhuman and degrading by the Committee for the Prevention of Torture and Inhuman or Degrading Treatment or Punishment (CPT).¹¹⁶ Cases in which similarly painful or simply humiliating practices (including, for example, prolonged courses of heavy sedation) have been approved by doctors, but not been subjected to the scrutiny of bodies such as the CPT, are likely to prove more problematic.

Home

Without appropriate support, or in inappropriate surroundings, disabled people may live in what would generally be regarded as “degrading” conditions even though they are within their own homes. Some examples of the types of difficulties which they may face are provided by Rowena Daw’s report for the Disability Rights Commission.¹¹⁷ In cases such as *A v UK*¹¹⁸ and *Z v UK*,¹¹⁹ it was held that States have a positive obligation to take steps to protect children living in their own homes who are at risk of suffering inhuman or degrading treatment at the hands of people not connected with the State (in those cases parents or step-parents). In both these cases the children concerned were at risk of experiencing inhuman and degrading treatment through violence although, in *Z*, they were also at risk of experiencing it through neglect. There is little doubt that a similar duty would exist where people become vulnerable to the risk of such harm at the hands of others because of disability rather than childhood. There is, on the other hand, some uncertainty as to whether Article 3 imposes any obligation on signatory States to take steps to protect disabled people from the isolation, humiliation, degradation and injury which sometimes result from living in an inaccessible environment or without adequate support. This issue has been considered in two recent English cases which appear to have adopted quite different approaches.

In *R v Enfield LBC, ex p Bernard*,¹²⁰ a restrictive approach to Article 3 was adopted. From June 2000 until October 2002 Mrs Bernard had lived in “deplorable conditions”,¹²¹ with her husband and six children, in a house which had been provided by the Local Authority and which was not accessible to her. Sullivan J took the view that the delay and inefficiency of the Council (which amounted to a breach of its statutory duty to provide suitable accommodation under section 21(1)(a) of the National Assistance Act 1948) had unnecessarily lengthened her ordeal in that house by some twenty months.¹²² Reports drawn up in September 2000 had informed the Authority of her situation and the urgent need to provide her with alternative accommodation, but no action had been taken.

Caged Beds” *Disability Tribune* (London: Disability Awareness in Action, December 2003/January 2004) reporting on the use of such beds in Slovenia, Hungary, Slovakia, and the Czech Republic.

¹¹⁶ CPT/Inf (1998) 12 para.39; CPT/Inf (2001) 2 para.8.

¹¹⁷ *The Impact of the Human Rights Act on Disabled People* (n.92 above), 54-55.

¹¹⁸ (1998) 100/1997/884/1096 (23 September 1998).

¹¹⁹ (2002) 34 EHRR 97.

¹²⁰ [2002] EWHC (Admin) 2282.

¹²¹ *ibid.*, at paras.[29] and [30].

¹²² *ibid.*, at para.[31].

Mrs Bernard's electric wheelchair could not be taken into the house and, as a result, she spent most of her waking time sitting in a shower chair which caused her pain. She was unable to reach the bathroom without help from her husband, who was not always available. Consequently, she soiled herself several times each day: a situation which, according to her husband, she found "depressing", "demeaning" and "very humiliating".¹²³ She attempted to reduce this problem by drinking less, which exacerbated her diabetes. She could not reach or use the kitchen and was therefore unable to cook for herself and her family. She was unable to answer the door or to leave the house independently. She was confined to the lounge which, at night, became a bedroom for herself and her husband and their two youngest children. Though the judgement does not contain information about her falls and injuries after the first few months in the house, it does indicate that, by September 2000, she had fallen four times, injuring her breast and her thumb. Not surprisingly, the social worker's report compiled in that month drew attention to the fact that:

"Mrs Bernard feels very unsafe at home. Occasionally she is left alone with the two youngest children when her husband has to collect the other children from school. At these times she feels particularly concerned should there be an accident/emergency she could not raise an alarm or help the children. . . Generally feels very vulnerable in her situation."¹²⁴

Although Mrs Bernard's case succeeded under Article 8, and indeed constituted a landmark development on the issue of damages under section 8 of the Human Rights Act 1998, it did not succeed under Article 3. Sullivan J acknowledged that the conditions in which she had lived were deplorable and that they had deprived her of her ability "to operate . . . as a person in her own right" and of her "dignity as a human being".¹²⁵ They were, in his view, conditions which were likely to have caused a less resilient person to have become mentally or physically ill.¹²⁶ Nevertheless, Sullivan J found that the minimum level of severity threshold had not been crossed.¹²⁷

This aspect of Sullivan J's decision is extremely disappointing. He explained it on the basis that cases such as *Price*, in which Article 3 was found to have been violated despite the lack of any intention to humiliate or debase, were concerned with the rights of prisoners and must be treated "with great caution outside the prison gates". Prisoners were deliberately subjected, against their will, to regimes ordained by prison authorities. Mrs Bernard, on the other hand, had been living in her own home, surrounded by her family, and the deplorable conditions which she had had to endure had not been "deliberately inflicted" upon her or the result of a "positive decision". They were, rather, the result of the Council's "failure to act" and of its "corporate neglect".¹²⁸

¹²³ *ibid.*, at para.[20].

¹²⁴ *ibid.*, at para.[7].

¹²⁵ *ibid.*, at para.[33].

¹²⁶ *ibid.*, at para.[46].

¹²⁷ *ibid.*, at para.[28].

¹²⁸ *ibid.*, at paras.[28]-[29].

Sullivan J's analysis of this issue is brief and also somewhat confusing. The confusion arises in part because he did not differentiate between two separate, though inter-related, questions: first, whether the suffering had reached the minimum Article 3 threshold of severity and, second, whether the State should be held responsible for that suffering.¹²⁹ While Sullivan J ruled that Mrs Bernard's suffering did not satisfy the threshold requirement, it is unclear whether his reasons for this observation relate primarily to this first issue or whether they, in fact, relate to the question of State responsibility for the suffering concerned (the second question).

In relation to the first issue (that of severity of suffering), the fact that it occurred in a home environment is undoubtedly relevant. The fact that suffering is experienced in an institution, such as a prison, into which the victim is forced to enter and over which they have no control, should undoubtedly be regarded as an exacerbating factor.¹³⁰ To this extent, Sullivan J's differentiation between *Price* and *Bernard* is entirely appropriate. His consideration of whether Mrs Bernard's suffering was sufficiently severe to cross the Article 3 threshold, however, should not have stopped here. To regard the absence of this factor as necessarily decisive would be to fail to acknowledge the extent of the degradation and humiliation which may be experienced by disabled people inside their own homes. Mrs Bernard, for instance, was effectively imprisoned in a shower chair in her living-room for over two years; her husband holding the key. When he was not there, she was trapped and frightened: unable to leave the house or to use the bathroom. She was required to undergo the indignity of soiling herself (and often the carpet) two or three times a day in the presence of her young children and suffered physical injuries as a result of futile attempts to reach the bathroom and avoid this indignity. Furthermore, as Sullivan J recognised, her ordeal was an extremely protracted one (lasting for twenty months longer than necessary) as opposed to the week long ordeal endured by Ms Price. It is arguable that, even though Mrs Bernard was not a prisoner as such, her ordeal amounted to a form of imprisonment. It is also certainly arguable that it involved "actual bodily injury or intense physical or mental suffering" or that it humiliated or debased her, "showing a lack of respect for, or diminishing . . . her human dignity or arouse[d] feelings of fear, anguish or inferiority capable of breaking [her] moral and physical resistance".¹³¹

It is a great pity that the second issue, that of state responsibility for suffering of the type endured by Mrs Bernard, was not considered more fully by Sullivan J. He indicated that Article 3 was not violated because the suffering had been caused by neglect rather than by a deliberate act. As mentioned above, it is not entirely clear whether this point was intended to address the issue of the severity of Mrs Bernard's suffering or whether it was intended to address the issue of state responsibility. Assuming it was directed at the

¹²⁹ See, e.g. *Gezer v Sec of State for the Home Dept* [2004] EWCA (Civ) 1730, para.[25] per Laws LJ for a clear identification of these two separate questions.

¹³⁰ For recognition of the importance of this distinction, see *R v DPP, ex p Pretty* [2002] 1 AC 800, para.8 per Lord Bingham and *Keenan v UK* (2001) 33 EHRR 38, para.90.

¹³¹ *Pretty v UK* (2002) 35 EHRR 1, para.52.

latter, it is an inadequate explanation for a refusal to hold the Council liable for an Article 3 violation.

In certain circumstances, as explained above, the failure of a public authority to take action may well violate Article 3. The *Bernard* case, being one in which the suffering was caused by circumstance rather than by violence, would certainly fall towards the outer edge of category (b) of Laws LJ's spectrum. Various features of the case, including the intensity of Mrs Bernard's suffering, its duration and the fact that it resulted from a grossly negligent failure to implement existing policy, would seem to demand a consideration of the breadth of *Z v UK*¹³² and the question of whether the State's failure to act was within the margin of judgment granted to it by Article 3 – a question which would require consideration of its impact on Mrs Bernard and the consequences of the finding of a violation on the general population. Such questions deserve careful scrutiny and it is to be hoped that, in any future similar case, they will receive it. It should perhaps be stressed that a finding of an Article 3 violation in cases such as *Bernard* need not oblige States to introduce a basic minimum level of provision for disabled people.¹³³ As in *Z*, what is at issue is the impact on a particular individual of the failure of a State to implement its accepted policy.

The second of the recent first instance cases to have considered Article 3 in the context of disabled people living in their own home is *R v East Sussex County Council, ex parte A and B*.¹³⁴ In this case a more generous approach than that in *Bernard* appears to have been adopted. Its likely impact is reduced, however, by the fact that its analysis of the Article is extremely brief, is very much obiter and contains no reference to the *Bernard* decision.

The East Sussex case concerned two sisters in their twenties, both of whom had profound physical impairments and learning difficulties. Because of the Council's "no lifting" policy (which forbade its staff from lifting people such as A and B manually), the sisters became virtual prisoners in their parents' home. The blanket "no lifting" policy, combined with the lack of mechanical lifting devices, meant that they were unable to access toilets and other facilities outside their home. The introduction of this policy was therefore challenged.

Though not required to decide whether Article 3 had been violated, Munby J indicated that in this type of case the Article might be engaged where, for instance, failing to lift A or B manually might result in them remaining "sitting in bodily waste or on the lavatory for hours, unable to be moved".¹³⁵ He drew attention to the fact that the standards required by Article 3 increased as social standards improved and observed that:

"The concept of human dignity may be the same as ever, but the practical standards which require to be met are not. Changes in social standards demand better provision for the disabled if their human dignity is not to be impaired."¹³⁶

¹³² [2002] 34 EHRR 97.

¹³³ See the text accompanying nn.23-26 above.

¹³⁴ [2003] EWHC (Admin) 167.

¹³⁵ *ibid.*, at para.[114].

¹³⁶ *ibid.*, at para.[98].

In the UK, then, there is currently some uncertainty as to how Article 3 should be applied in cases where disabled people, not living in prisons or other residential institutions, suffer humiliation, indignity and discomfort as a result of inadequate support. The two cases which have considered the issue are both first instance and point in different directions. If there is indeed to be no Article 3 violation in a case such as *Bernard*, it is virtually impossible to imagine any circumstances in which disabled people living in their own homes would be able to establish an Article 3 violation in the absence of a deliberate intent that they should suffer. The issue is an important one, particularly if disabled people are to be permitted, and even encouraged, to move out of institutions and into their own homes.¹³⁷ It requires much closer judicial scrutiny than it has received to date.

The Universality of Article 3

The question of whether Article 3 applies where the disputed treatment is inflicted on a person who is unaware of, or unable to appreciate, any consequent suffering or degradation has not yet directly confronted the ECtHR. In recent years, however, it has troubled UK courts on a number of occasions and resulted in conflicting decisions. The context in which this question has arisen is that of disputes about the withdrawal of artificial nutrition and hydration from patients in a “persistent vegetative state”.¹³⁸ Significantly, however, it does have relevance to, and implications for, people who are conscious and even active but who have mental impairments or illnesses which render them unable to appreciate that what is happening to them would conventionally be regarded as humiliating or degrading.

In the pre-Human Rights Act case of *Airedale NHS Trust v Bland*,¹³⁹ the House of Lords ruled that it would be lawful for doctors to withdraw treatment (which was held to include artificial nutrition and hydration) from a patient if the treatment in question was considered by a responsible body of medical opinion to be no longer in the best interests of that patient. Tony Bland, who had been in a persistent vegetative state for three years after having been trampled on at the Hillsborough football stadium, was considered to have effectively no interests at all. It was therefore lawful to remove the tubes which supplied him with food and water even though, after a period of about two weeks, this would inevitably result in his death.

The *Bland* decision has generated much debate. Concerns expressed by commentators relate to the influence of medical judgments that some lives

¹³⁷ Cabinet Office Strategy Unit *Improving the Life Chances of Disabled People* (n.87 above), chap.4. See also Article 15 of the UN Draft Comprehensive and Integral International Convention to Protect and Promote the Rights and Dignity of Persons with Disabilities, available at www.un.org/esa/socdev/enable/rights/ahcwgreporta19.htm (last visited on 30 September 2005) which would confer on disabled people the right to choose to live in their communities with appropriate support and prevent States from requiring them to live in institutions.

¹³⁸ Although this term is now in general use, it is unpopular with many doctors and with patients’ associations because of its tendency to equate the people concerned with vegetables. For further discussion, see A. Grubb, P. Walsh and P. Lambe, “Reporting on the Persistent Vegetative State in Europe” [1998] *Med LR* 161, 173.

¹³⁹ [1993] AC 789.

are not worth living in decisions to end life;¹⁴⁰ to the possibility of misdiagnoses of a persistent vegetative state and mistaken prognoses as to the likelihood of recovery;¹⁴¹ and to the possible extension of *Bland* beyond people in persistent vegetative states to people with other forms of physical or mental impairment.¹⁴² Although these issues are of supreme relevance to disabled people, they fall outside the ambit of Article 3 and therefore of this discussion.

The relevance of *Bland* to this article lies in the argument, to which it has given rise, that the withdrawal of food and water from a person may constitute inhuman or degrading treatment contrary to Article 3. This argument was raised and rejected in *NHS Trust A v Mrs M, NHS Trust B v Mrs H*.¹⁴³ Butler-Sloss P, after referring to the need for “degrading treatment” to arouse feelings of fear, anguish or inferiority in the victim, declared herself to be “satisfied that Article 3 requires the victim to be aware of the inhuman and degrading treatment which he or she is experiencing or at least to be in a state of physical or mental suffering”.¹⁴⁴ Because the two people concerned in that case were in persistent vegetative states, they had “no feelings and no comprehension of the treatment accorded to [them]”¹⁴⁵ and Article 3 could therefore not apply.

This conclusion is alarming.¹⁴⁶ It raises the difficult question of exactly how much consciousness or awareness a person must have in order to enjoy the protection of Article 3. Would its protection extend, for instance, to a person who is fully conscious but who, because of a mental impairment, does not interpret what is happening to them as a humiliation or degradation?

Butler-Sloss P’s ruling places disabled people who are unable to appreciate the humiliating, degrading or even painful consequences of particular treatment (whether because they are in a persistent vegetative state or because they have some lesser degree of mental impairment), completely beyond the protective scope of Article 3. That Article would not be infringed if, for instance, an insensate or unaware patient were used as a

¹⁴⁰ See, e.g. J. Keown, “Restoring Moral and Intellectual Shape to the Law After *Bland*” (1997) 113 LQR 481, especially at 492-496; J.M. Finnis, “*Bland*: Crossing the Rubicon” (1993) 109 LQR 329.

¹⁴¹ See, e.g. J. Keown, “Restoring Moral and Intellectual Shape to the Law After *Bland*” (1997) 113 LQR 481, 498.

¹⁴² See, e.g. G.T. Laurie and J.K. Mason “Negative Treatment of Vulnerable Patients: Euthanasia by any other Name?” [2000] *Juridical Review* 159; J. Keown, “Life and Death in Dublin” [1996] CLJ 6. For an example of a decision which has authorised the removal of artificial nutrition and hydration from patients who have some degree of cognitive functioning, see *In the Matter of a Ward of Court* [1995] 2 ILRM 410. See, for discussion of ‘slippery slope’ arguments, S.W. Smith, “Evidence for the Practical Slippery Slope in the Debate on Physician-Assisted Suicide and Euthanasia” [2005] *Med LR* 174.

¹⁴³ (2001) 58 BMLR 87. The same conclusion was reached, *obiter*, in *A National Health Service Trust v D* [2000] 2 FLR 677 *per* Cazalet J.

¹⁴⁴ *ibid.*, at para.[49].

¹⁴⁵ *ibid.*

¹⁴⁶ For cogent criticism of the reasoning, see A. Maclean, “Crossing the Rubicon on the Human Rights Ferry” [2001] 64 MLR 775, 789-792.

sideboard¹⁴⁷ or if they were “treated as a sex object or thrown, living, into the hospital rubbish”,¹⁴⁸ It is a conclusion which does not sit easily with the observation of the ECtHR in *Herczegfalvy* that the “position of inferiority and powerlessness which is typical of patients confined in psychiatric hospitals calls for increased vigilance in reviewing whether the Convention has been complied with”.¹⁴⁹ It is a conclusion which fails to acknowledge that “the degradation of an incapacitated person shames us all even if that person is unable to appreciate it”.¹⁵⁰ It is a conclusion which refuses to accept that human rights are universal: to be conferred on every human being simply by virtue of their humanity.

The issue was raised again at first instance in *R v General Medical Council, ex p Burke*¹⁵¹ where it was described by Munby J as “a point of absolutely fundamental importance”.¹⁵² He stressed the significance of the Convention’s role in protecting the human dignity of disabled people, whom he considered to be particularly vulnerable to abuse,¹⁵³ and drew support from observations made in *Keenan v UK*¹⁵⁴ and *Pretty v UK*¹⁵⁵ to conclude that the approach of Butler-Sloss P had been “wrong”.¹⁵⁶ In *Keenan*, the ECtHR had noted that, though severity of suffering had been a significant issue in many of the Article 3 cases, “there are circumstances where proof of the actual effect on the person may not be a major factor”, such as in the case of a mentally-ill person who is incapable of “pointing to any specific ill-effects”.¹⁵⁷ In *Pretty*, degrading treatment had been described as treatment which either “humiliates or debases an individual, showing a lack of respect for, or diminishing, his or her human dignity” or which arouses in him or her feelings of fear, anguish or inferiority.¹⁵⁸

Munby J, then, ruled that treatment inflicted on a person unable to appreciate suffering may breach Article 3. The victim’s awareness and experience of suffering were relevant, but not essential, factors in the determination of whether the Article had been infringed. His reasoning on this point is clearly set out in the following passage:

“In my judgment treatment is capable of being “degrading” within the meaning of Article 3, whether or not it arouses feelings of fear, anguish or inferiority in the victim. It is

¹⁴⁷ A scenario drawn from J. Keown, “Restoring Moral and Intellectual Shape to the Law After *Bland*” (1997) 113 LQR 481, 494.

¹⁴⁸ Scenarios drawn from J.M. Finnis, “*Bland*: Crossing the Rubicon” (1993) 109 LQR 329, 336. See also A. Maclean, “Crossing the Rubicon on the Human Rights Ferry” [2001] 64 MLR 775, 791, for discussion of a US case concerning the rape of a woman in a persistent vegetative state.

¹⁴⁹ (1992) 15 EHRR 437, para.82.

¹⁵⁰ *R v Broadmoor Special Hospital Authority, ex p Wilkinson* [2002] 1 WLR 419, para.[79] per Hale LJ.

¹⁵¹ [2004] EWHC (Admin) 1879.

¹⁵² *ibid.*, at para.[146].

¹⁵³ See, e.g. *ibid.*, at paras.[67]-[72].

¹⁵⁴ [2001] 33 EHRR 913.

¹⁵⁵ [2002] 35 EHRR 1.

¹⁵⁶ [2004] EWHC (Admin) 1879, para.[146].

¹⁵⁷ [2001] 33 EHRR 913, para.[113].

¹⁵⁸ [2002] 35 EHRR 1, para.[52].

enough if judged by the standard of right-thinking bystanders . . . it would be viewed as humiliating or debasing the victim, showing a lack of respect for, or diminishing, his or her human dignity.”¹⁵⁹

This passage, it is suggested, merits a prominent position in the jurisprudence of Article 3. Regrettably, however, this does not appear to be its destiny. The Court of Appeal in *Burke*¹⁶⁰ did not address the issue of the applicability of Article 3 to people lacking awareness of suffering or degradation; clearly considering it to be a matter on which the Court was not required to provide any direction. Their Lordships stated that it would be ‘inappropriate’ to allow Munby J’s judgement, which was severely criticised for its detailed consideration of issues not directly raised by the facts of the case,¹⁶¹ to be “seized on and dissected by lawyers seeking supportive material for future cases”.¹⁶² They warned that:

“Although we have said that a great deal in the body of the judgment is uncontroversial, we counsel strongly against selective use of Munby J’s judgment in future cases.”¹⁶³

It thus appears that the approach of Butler-Sloss P in *NHS Trust A v Mrs M, NHS Trust B v Mrs H*¹⁶⁴ represents the current law. According to it, the withdrawal of food and water from (or the infliction of any form of treatment or insult upon) people in persistent vegetative states could never amount to an Article 3 violation. Nevertheless, it is worth considering briefly how the withdrawal of food and water from such people would be regarded were Munby J’s approach to the applicability of Article 3 to be adopted instead.

In *NHS Trust A v Mrs M, NHS Trust B v Mrs H*,¹⁶⁵ Butler-Sloss P observed that, should Article 3 be applicable to the two patients with whom she was concerned, it would not be violated by the termination of artificial nutrition and hydration. She referred to the “therapeutic necessity” test laid down in *Herczegfalvy* and was satisfied that, in the case before her, the proposed withdrawal of treatment had been “thoroughly and anxiously considered by a number of experts in the field” and was “in accordance with the practice of a responsible body of medical opinion”.¹⁶⁶ In *Burke*, Munby J expressed broad agreement with this approach¹⁶⁷ but stressed that conceptions of what was dignified or undignified would vary from individual to individual and drew attention to the overriding importance of the right to personal autonomy protected by Article 8.¹⁶⁸ In assessing whether the withdrawal of artificial nutrition and hydration from an insensate person would infringe Article 3, much would therefore depend on the terms of any advanced directive they

¹⁵⁹ [2004] EWHC (Admin) 1879, para.[149].

¹⁶⁰ [2005] EWCA (Civ) 1003.

¹⁶¹ *ibid.*, at paras.[19]-[21].

¹⁶² *ibid.*, at para.[24].

¹⁶³ *ibid.*

¹⁶⁴ (2001) 58 BMLR 87.

¹⁶⁵ *ibid.*

¹⁶⁶ *ibid.* at para.49.

¹⁶⁷ [2004] EWHC (Admin) 1879, para.[136].

¹⁶⁸ *ibid.*, at paras.[66] and [127].

had made and their beliefs and values, as well as on their medical condition at the time in question.¹⁶⁹

The applicability of Article 3 to people unable to comprehend the humiliation or suffering which the treatment in question would otherwise cause them is, as Munby J recognised, an issue of “absolutely fundamental importance”. It is, of course, of importance to people who lack this capacity (or who may do so in the future) and to their families. It is also of importance to society as a whole and to the nature of the Convention’s role. If human rights are to be in any meaningful sense “universal”, it is vital that Munby J’s approach is adopted by higher courts in the UK and, of course, by the ECtHR, in preference to that outlined by Butler-Sloss P.

Conclusion

Article 3 plays an essential role in the protection of the humanity and dignity of all (or, on the view of Butler-Sloss P in *NHS Trust A v Mrs M, NHS Trust B v Mrs H*,¹⁷⁰ nearly all) citizens in signatory States. It is, of course, not the only ECHR article to offer protection and, in many cases, it may be possible to establish a violation of Article 8 in circumstances which would fall short of the severity required by Article 3. It is important, however, that judges and advocates resist any temptation to avoid consideration of Article 3 in cases where it is clear that redress will be available through other means. Article 3, in marking out the most unacceptable instances of State-generated degradation and humiliation, has a powerful declaratory function which should not be overlooked.

It has been said that: “Disability challenges society to live up to its values and sets [human rights] law the task of engineering positive change”.¹⁷¹ Many of the Article 3 cases discussed here have set about this task with genuine commitment. Nevertheless, in some important respects the Article 3 jurisprudence lends what is, at best, only equivocal support. In particular, on the most fundamental issue of all – the issue of whether certain categories of disabled people are altogether excluded from the protective scope of Article 3 – there is currently uncertainty.

A number of decisions relating to disabled prisoners demonstrate the potential of Article 3 for disabled people. Its focus on the suffering of the particular individual obliges States to respond to the particular circumstances of a disabled person and, on occasion, to treat them differently from others. Such recognition of, and responsiveness to, the needs of a particular individual is crucial to any meaningful notion of respect or dignity.¹⁷²

¹⁶⁹ *ibid.*, at para.[176].

¹⁷⁰ See the discussion in the previous section of this article.

¹⁷¹ G. Quinn and T. Degener (eds.), *Human Rights and Disability - the current use and future potential of United Nations human rights instruments in the context of disability* (n.3 above), para.1.0.

¹⁷² For further discussion of the concept of dignity, see D. Beyleveld and R. Brownsword, “Human Dignity, Human Rights and Human Genetics” [1998] MLR 661; D. Feldman, “Human Dignity as a Legal Value – Part I” [1999] PL 682; and G. Quinn and T. Degener (eds.), *Human Rights and Disability - the current use and future potential of United Nations human rights instruments in the context of disability* (n.3 above), chap. 1.

These prison cases have direct relevance to the treatment of disabled people in other forms of residential institution where a high degree of control is exercised over all aspects of their lives. It is disappointing, however, that there have not been more cases challenging the treatment of disabled people in such institutions.¹⁷³ The dearth of cases should certainly not be regarded as proof that all is well. It is to be hoped that improvements will emerge from the work of the newly created Commission for Social Care Inspection.¹⁷⁴ The raising of awareness of human rights amongst those living and working in such institutions should also be a priority for the Commission for Equality and Human Rights (which is likely to be established shortly and charged with a duty to promote awareness of human rights).¹⁷⁵

Outside the institutional context, the UK disability cases are currently inconclusive. Though Article 3 has received painstaking analysis in the context of asylum seekers, its treatment in the disability context has been somewhat peremptory. There is, it is suggested, much scope for the development of Article 3 jurisprudence and protection in this area. It is sometimes said, with good reason, that disabled people have traditionally been “invisible” to human rights law;¹⁷⁶ an invisibility which, as some commentators have speculated, may arise from a belief that disabled people are not yet “ripe for freedom”¹⁷⁷ or from a tendency to view them as somehow “other” and therefore not truly human.¹⁷⁸ Although cases such as *Price* demonstrate a welcome recognition of their existence, their presence on the human rights stage is still little more than shadowy. If society is to demonstrate true commitment to the dignity and value of disabled people, it must allow Article 3 to play a role outside the prison gates and to reach into the unimaginable and unnecessary suffering and degradation experienced by people such as Mrs Bernard. Until this happens, aspirations to meaningful inclusion and participation will remain little more than unrealisable dreams for many disabled people.

¹⁷³ See further, L. Clements and J. Read “The Dog that Didn’t Bark: The Issue of Access to Rights under the European Convention on Human Rights by Disabled People” (n.97 above).

¹⁷⁴ Set up in April 2004 by Part 2 of the Health and Social Care (Community Health and Standards) Act 2003 and charged with the inspection and regulation of social care provision.

¹⁷⁵ Equality Bill 2005 (introduced into the House of Lords on 18 May 2005). See also DTI, *Fairness for All: A New Commission for Equality and Human Rights* (London: Stationary Office, 12 May 2004); and DTI, *Government Response to the Consultation on Fairness for All: A New Commission for Equality and Human Rights* (London: Stationary Office, 18 November 2004).

¹⁷⁶ See, e.g. G. Quinn and T. Degener (eds.), *Human Rights and Disability - the current use and future potential of United Nations human rights instruments in the context of disability* (n.3 above); and A. Hendriks, “Disabled Persons and Their Right to Equal Treatment” in J.M. Mann, S. Gruskin, M.A. Grodin and G.J. Annas (eds.), *Health and Human Rights* (London: Routledge, 1999).

¹⁷⁷ L. Clements and J. Read, “The Dog that Didn’t Bark: The Issue of Access to Rights under the European Convention on Human Rights by Disabled People” (n.97 above), 25.

¹⁷⁸ *ibid.*, at 25-26. See also, A. Lawson, “The Human Rights Act 1998 and Disabled People: A Right to be Human?” in C. Harvey, (ed.), *Human Rights in the Community* (Oxford: Hart Publishing, 2005).

In summary, then, disabled people continue to experience difficulties in accessing the protection of Article 3. For those living in institutional care, these difficulties are largely practical in nature; difficulties which make it extremely unlikely that Article 3 abuses will be tackled through the courts. It is therefore important that relevant authorities work to prevent the occurrence of abuse and suffering in these institutions as is required, indeed, by the positive obligations of Article 3. For those living in their own homes, the difficulties are more jurisprudential in nature. The courts have, to date, failed to develop and apply the principles of Article 3 to the suffering of disabled people in the non-institutional context. Finally, it must not be forgotten that over the entire subject there still hovers the dark cloud of an interpretation of Article 3 which would deny its protection to the most vulnerable of disabled people; an interpretation which has been ably challenged by Munby J but which has not yet been unequivocally dispelled. It is fitting to conclude with the words of Baroness Hale which, though referring to the ECHR as a whole, are particularly relevant to Article 3:

“We need to be able to use it to promote respect for the inherent dignity of all human beings but especially those who are most vulnerable to having that dignity ignored.”¹⁷⁹

¹⁷⁹ “What Can the Human Rights Act do for my Mental Health?”, Paul Sieghart Memorial Lecture London: British Institute of Human Rights, July 2004.

THE SIGNIFICANCE OF CORPORATE COLLECTIVE POWER TO NEW LABOUR'S PROJECT: MARKET VALUES, CORPORATE REGULATION AND THE POSSIBILITIES FOR A PROGRESSIVE POLITICS

Gary Wilson, *School of Law, University of Keele**

"The pure and sinless did not exist, or else died unnoticed and with no obituary. The idea seemed persuasive. Those who merited obituaries usually achieved things, fought for their ideals, and when locked in battle, it wasn't easy to remain entirely honest and upright. Today's battles were all for material gain, anyway. The crazy idealist was extinct- survived by the crazy pragmatist . . ."

A. Kurkov, *Death and the Penguin* (2003) 61

Introduction

This paper seeks to sketch the historical development of the political economy of the corporation in the United Kingdom from the onset of modernity to the beginning of the twenty-first century, in order to contextualise the significance of corporate power to the politics of New Labour and to the broad possibilities offered by the major review of company law legislation commenced in 1998.¹ It will be argued that the way in which the corporate entity is theorised and regulated is not simply a subject for abstruse academic debate or for resolution at a technical level by expert committee (though, for sure, both do have an important role) but is rather a matter of central import to the broader trajectory of modern industrial societies, both in terms of the appropriate regulation of collective power and the generation of wider values beyond those of market individualism. Accordingly, it thus posits a key issue for mainstream progressive politics.

In order to substantiate this claim it will be necessary to interleave an analysis of the pertinent constructs underpinning New Labour's political programme (notably globalisation, post-Fordism, and the recent work of Anthony Giddens² on the Third Way and progressive politics); the different

* The author would like to thank Michael Cardwell, Matthew Weait and Sarah Wilson for their encouragement and comments. This article is written as at 1 August 2005: the Company Law Reform Bill has subsequently been put before Parliament.

¹ Following the publication of *Modern Company Law for a Modern Economy: Final Report* (2001) by the Company Law Review Steering Group in June 2001, the Government has issued two White Papers setting out proposals for consultation on a new Companies Act: see *Modernising Company Law* (July 2002; Cm 5553) and *Company Law Reform* (March 2005; Cm 6456).

² The principal works are as follows: *Beyond Left and Right: The Future of Radical Politics* (1994); *The Third Way: The Renewal of Social Democracy* (1998); *The*

approaches taken to the legal theorisation of the corporate entity; and the present extensive corporate law reform programme in the United Kingdom.³ Particular emphasis will be given to attempting to capture the way in which consideration of the collective power inherent in corporations would appear to be notable largely by its absence in both corporate law theory and New Labour's overall worldview. It will be suggested that the relative invisibility of this collective power has potentially profound ramifications for the possibilities inherent in a progressive politics (and thus for society as a whole) given the juxtaposition of New Labour's well known desire to be business friendly⁴ and the increasing emphasis placed by New Labour politicians upon the role of a distinctive set of values as the marker of a progressive agenda.⁵

The quote from Andrey Kurkov's *Death and the Penguin* at the head of this article might indeed be thought to be prescient (the novel was first published in 1996) when applied to the common critiques of New Labour's general pro-business/pro-market philosophy and period in office, especially from those on the traditional Left. At the same time it would not seem too far from the mark to characterise the Labour Party's own trajectory over the last 25 years as shifting from 'crazy idealism' to 'crazy pragmatism'. The Party's initial reaction to Thatcherism was to maintain a relatively ideologically pure socialist position that placed the party in the political wilderness for over a decade (the manifesto for the 1983 general election⁶ was famously dubbed 'the longest suicide note in history'). As a direct response to this Blair's internal party reforms of the mid-1990's⁷ led to the creation of New Labour and the development of a rather murky set of social democratic principles which, whilst leading to repeated electoral success (an unprecedented third term was achieved in May 2005), sacrificed many fundamental Labour Party

Third Way and its Critics (2000); and *Where Now for New Labour?* (2002). Edited collections comprise: *The Global Third Way Debate* (2001); *The Progressive Manifesto: New Ideas for the Centre-Left* (2003); with W. Hutton, *Global Capitalism* (2000); and with P. Diamond, *The New Egalitarianism* (2005).

³ Insolvency law has also been extensively reformed in line with the New Labour emphasis on fostering entrepreneurship through seeking to reduce the stigma of failure in personal insolvency and to develop an enhanced rescue culture in corporate insolvency: see Insolvency Act 2000, Enterprise Act 2002 and G. Wilson and S. Wilson, 'Responsible Risk-Takers: Notions of Directorial Responsibility-Past, Present and Future' (2001) 1 *J. Corp. Law Studies* 211.

⁴ At each general election since 1997 the party has produced a detailed separate business manifesto. Blair's speech (28 April 2005) launching the 2005 business manifesto was headed 'Labour is the Party of Business.'

⁵ See, e.g. T. Blair, 'Progressive Values', (Spring 2005) 4.1 *Progressive Politics* 57-63. This journal is sponsored by Policy Network, an international independent think tank for the development of progressive ideas of the centre-left chaired by Peter Mandelson.

⁶ *A New Hope for Britain*: policies included Keynesian macro-economics, re-nationalization of privatized industries and withdrawal from the European Community.

⁷ It is important to note that Blair was able to build upon work undertaken by the two previous Labour Party leaders, Neil Kinnock and John Smith, but the New Labour soubriquet that has been used since shortly after Blair's election as leader in 1994 was clearly intended to signify a fundamental re-configuration of party policy.

precepts⁸ in order, according to the modernisers, to regain relevance in the profoundly changed circumstances of a globalized world; and according to the critics, to facilitate the pragmatic pursuit of power.⁹

The burden of this article will be to argue that there is substance to the New Labour programme (whether labelled as the Third Way or reincarnated under its current appellation progressive politics) but that to date the nature of corporate collective power has been both inconsistently recognised and inadequately theorised in this context. This has created a serious lacuna, which is particularly important given the prominence of the business sector in the present United Kingdom political firmament and the stress that is laid upon values by those professing a progressive politics. As was noted by a former Secretary of State for Trade and Industry, Stephen Byers MP (when promoting discussion about corporate governance as part of the independent company law review process), the nature of the prospective reform of United Kingdom corporate law is thus of crucial significance well beyond the narrow parameters of company law itself:

“The reason this debate is important is that the way that companies are allowed to operate defines the nature of the market economy that we live in. The key to shaping the market in ways that achieve our twin objectives of efficiency and social justice lie in the framework of rules within which companies do business and make a profit. So company law and corporate governance are at the heart of the debate about the kind of society we want and the nature of our economy.”¹⁰

It is interesting to note that, whilst Byers’ initial sentence is addressing the market economy, there is subsequently clear recognition of the pervasive (indeed, to those of a critical disposition, corrosive) nature of the broad highly individualist values and *modus operandi* originating in the market domain with respect to society more generally.¹¹ Given that New Labour’s preference for market solutions has also been vigorously extolled within the sphere of the European Union, the United Kingdom is one of the key sponsors of the Lisbon Agenda,¹² it is not surprising that this issue has

⁸ Most notably, at least in symbolic terms, by the repeal of Clause 4 of the Party’s constitution (re common ownership of the means of production via nationalization) at the October 1994 party conference.

⁹ See T. Blair, *The Third Way: New Politics for a New Century* (1998). Blair’s frank admission of the pragmatic nature of the project, ‘what matters is what works to give effect to our values’ (*ibid.*, p.4) (together in later years with an undue reliance on presentation or ‘spin’) has provided much grist to the critics’ mill: see, e.g. the articles in the special issue of *Marxism Today* (December 1998).

¹⁰ Speech made at the TUC / IPPR seminar on corporate governance (7 June 2000).

¹¹ Byers’ subsequent resignation as a Minister over the Railtrack administration in May 2002 and recent admission (in an ultimately unsuccessful High Court action brought against him for public misfeasance by aggrieved Railtrack shareholders) that he told an untruth to Parliament, are in this light especially unfortunate from a governance perspective. For a discussion of some of the issues involved, together with an analysis the nature of Railtrack’s successor company Network Rail, see L. Whitehouse, ‘Railtrack is Dead – Long Live Network Rail? Nationalization under the Third Way’ (2003) 30 *J. of Law & Soc* 217.

¹² The Agenda was adopted at the Lisbon European Council in 2000 with the principal aim of making the European Union the most dynamic and competitive

resonated badly with more social market orientated members of the Union, whose outlook was concisely put by the former French Prime Minister, and socialist, Lionel Jospin's remark, 'yes to a market economy, no to a market society'.¹³ The article will conclude with an examination of the extent to which the current corporate law reform proposals clearly address the collective power of corporations and are thus congruent with the development of a genuinely progressive politics as opposed to the adoption of a defensive position essentially framed by neo-liberalism and resulting in a society thoroughly infused by the values and culture of market individualism.

The Rise and Demise of Collectivism

In order to progress it is first necessary to attempt to chart the import of the notion of collective power, and corporate collective power in particular, as these ideas are central to the issues this article seeks to explore. Further, as power may for current purposes be said to be located in the temporal relations to be found within the complex matrices of agents (consisting of both institutional entities and individual persons) and structural conditions that compose society, the overall dimensions of collective power will only be rendered amenable to illumination, albeit feeble, through a broad investigation of pertinent changes in society over time.¹⁴ Hence, it will be argued that collective power is fully intelligible only in the context of the particular stage of modernity under examination and that the ambit of such power (and thus its precise significance) will be subject to modulation over time as the rationalities of modernity play out and thereby re-shape its very fabric.¹⁵ At a more concrete level the analysis will be pursued through the ideas of combination and especially of collectivism (which became the dominant mode of thinking in the latter part of the nineteenth century).¹⁶ In

knowledge-based economy in the world, capable of sustainable economic growth with more and better jobs and greater social cohesion, and respect for the environment.

¹³ See further, L. Jospin, *Modern Socialism* (1999). Like Blair, Jospin also places his justification for social democracy upon values, but the emphasis is quite different and is explicitly orientated around the regulation of the so-called 'natural' capitalist model (*ibid.*, p.10).

¹⁴ That such developments are believed to occur should not be taken as an endorsement of a Whiggish or Hegelian approach to history as a progressive unfolding of events.

¹⁵ There is not space here to pursue the thorny conceptual issues raised by the notion of post-modernism: for a brief discussion see K. Kumar, *From Post-Industrial to Post-Modern Society* (1995) 137-142, 173-183, and D. Harvey, *The Condition of Postmodernity* (1990) chap.6. The stance that will be taken in this article is that the current environment is one variously described as high modernity (Anthony Giddens), reflexive or second modernity (Ulrich Beck), or liquid modernity (Zygmunt Bauman); and which presents certain new conditions (which only Bauman would be likely to describe formally as post-modern) of necessity entailing a critical interrogation of its own rationalities.

¹⁶ Combination remained the dominant term in the United States but in the context of corporate consolidations such as the Standard Oil Trust, industrial concentration in the United Kingdom in general tended to occur significantly later: see respectively, M. J. Sklar, *The Corporate Reconstruction of American Capitalism*

line with the broader argument herein it is no coincidence that both of these concepts are more usually associated with the rise of organised labour for the collective nature of corporate power was seemingly early suppressed by the rise of a strong doctrine of separate corporate personality.¹⁷

For the celebrated constitutional lawyer Albert Dicey the growth of collectivism (admittedly largely focused around the recognition of trades unions and the enactment of social legislation) was to provide the key to his analysis of the relationship between the course of English thought and the development of English law during the nineteenth century and early twentieth centuries.¹⁸ However, despite the centrality of the concept of collectivism to his deliberations Dicey was at pains to confess that a precise definition of the term was not possible given its relative novelty and inherent ambiguity.¹⁹ Nevertheless, his stated preference was that collectivism consisted of two moments: a negative rebuttal of *laissez-faire* individualism as the basis for sound legislation and a positive endorsement of State intervention in order to benefit the mass of the population, even if this involved restrictions upon individual liberties.²⁰ Unfortunately, from the point of view of clarity, the second element is one that Dicey conflated with socialism, even though he expressly rejected any application of the concomitant economic precepts to his concept of collectivism, instead reiterating his own compass for the term, whereby it was to be ‘... used as a convenient antithesis to individualism in the field of legislation.’²¹

As Dicey himself acknowledged, the above definition was placed at an unhelpfully high degree of abstraction and for the purposes of this essay much may be gleaned from a close analysis of his application of the principle to concrete legislative situations. It is at this level of exposition that power relations are rendered discernible, though it should be borne in mind that Dicey was not directing his thoughts towards making different sites of collective power legible. Nevertheless, a rare explicit reference to collective power is made where the powerlessness of an individual worker against a corporation is advanced as part of the rationale for the emergent preference for collective action in English society.²² Significantly, however, the remainder of the discussion on this issue focuses around the legislative

1890-1916 (1988) and L. Hannah, *The Rise of the Corporate Economy* (2nd ed., 1983).

¹⁷ Until the mid-nineteenth century a company was referred to in the plural reflecting the combination of its shareholders but this linguistic practice changed to the singular when shares became fully autonomous tradeable property. The adoption of a robust legal concept of separate corporate personality from the 1860's is thus best seen as a reflection of this reification of the joint stock fund into money capital, see P. Ireland, 'Capitalism Without the Capitalist: the Joint Stock Company Share and the Emergence of the Modern Doctrine of Separate Corporate Personality' (1996) 17 *Journal of Legal History* 41.

¹⁸ A. V. Dicey, *Lectures on the Relation Between Law & Public Opinion in England During the Nineteenth Century* (2nd ed., 1962; first published 1905). Dicey set out three broad periods of development: Old Toryism/ Legislative Quiescence (1800-1830); Benthamism/ Individualism (1825-1870); and Collectivism (1865-1900).

¹⁹ *ibid.*, p.64, n.1, and p.67.

²⁰ *ibid.*, p.259.

²¹ *ibid.*, p.64, n.1.

²² *ibid.*, p.266.

mechanisms employed in the operationalization of the collective power of workers through the legitimation of trades unions²³ and no further consideration is given to the counterpoising power of the company.

This silence in relation to corporate collective power is despite the fact that Dicey had earlier deliberated upon the characteristics of modern commerce in the course of his discussion of the growth of collectivism.²⁴ Although the relevant section of the text is perhaps not as focused as now might be desired in the light of modern developments in company law,²⁵ Dicey's main conclusion was that through legislative development, ' . . . combination has gradually become the soul of modern commercial systems'.²⁶ Further, and in accordance with his general typology of collectivism, this 'revolution' in thought (to use Dicey's own term) had encouraged the State to interfere (on a constant basis) with individuals' private property rights in furtherance of business ends²⁷ and had also resulted in a tendency to celebrate the advantages of collective action and thereby to diminish the importance of individual endeavour.²⁸ It would seem that Dicey (perhaps unwittingly) had thus introduced another strand of meaning within his concept of collectivism by which it refers to the agency and collective power of a corporate body as a result of the combination of its shareholders and its legal and organisational structure.²⁹

²³ Dicey believed that legislative impulses leading to the enactment of Combination Acts in 1800, 1825 and 1875 exemplified his three stages of development. See *ibid.*, pp.271-73 and n.18 above.

²⁴ *ibid.*, pp.245-248.

²⁵ It would be unfair to blame Dicey for his lack of clarity of exposition regarding uses of the corporate form and the precise site of agency within a company as the 'one man' private company had only just been endorsed by the House of Lords in *Salomon v A. Salomon & Co Ltd* [1897] A.C. 22, and the relationship between the shareholders and the board of directors over management matters was still the subject of intense dispute until the decision, in the latter's favour, of the Court of Appeal in *Automatic Self-Cleansing Filter Syndicate Co Ltd v Cuninghame* [1906] 2 Ch. 34.

²⁶ Principally the Limited Liability Act 1855 and Joint Stock Companies Act 1862 as consolidated by the Companies Act 1862. For an extended discussion of nineteenth century developments in company law from a doctrinal and a socio-political view see respectively, B. C. Hunt, *The Development of the Business Corporation in England 1800-1867* (1936) and T. L. Alborn, *Conceiving Companies: Joint-stock Politics in Victorian England* (1998).

²⁷ Dicey contrasts a sole proprietor with a railway company having monopoly privileges and requiring compulsory purchase rights. The special facets of the latter are held to tie it closely to the state and to suggest public ownership. The relation between the corporation and state was a matter of considerable contemporary theoretical debate: see, e.g. F. Hallis, *Corporate Personality* (1930).

²⁸ Dicey, *op. cit.*, n.18, p.246, n.2 and p.247.

²⁹ See n.25 above. The nature of combination is not explored but the rise of managerialism due to the dispersed nature of personal shareholdings and the subsequent shift from personal to impersonal shareholding resulting in, 'a polyarchic structure of hegemonic financial intermediaries which participate in the controlling constellations of . . . major enterprises' were well in the future: see further, J. Scott, *Corporations, Classes and Capitalism* (2nd ed., 1985) (quotation at p.259).

History can make for surprising bedfellows and for Karl Marx, also commentating on the growth of joint-stock companies in the mid-nineteenth century (albeit from a very different political and theoretical perspective), there was no doubt that capital possessed this element of social power. Indeed, it was for this very reason that the joint-stock company represented the very negation of the private entrepreneurial capitalist³⁰ and thereby amounted to '[t]he abolition of the capitalist mode of production within the capitalist mode of production itself', being in essence '... private production unchecked by private ownership.'³¹

Marx was, of course, discussing the role of credit and joint-stock companies in the wider context of his analysis of society in the early stages of capitalist development. According to Marxist theory the motor of capitalism lay in its insatiable desire for accumulation through the medium of exchange on the market and its fuel lay in the commodification of the labour relationship between the owners and workers. The bourgeoisie were thus in a class war with the proletariat and this antagonism was the hallmark of such a thoroughgoing change in society that Marx and Engels discerned a significant discontinuity with the past:

“Constant revolutionizing of production, uninterrupted disturbance of all social conditions, everlasting uncertainty and agitation distinguish the bourgeois epoch from all earlier ones. All fixed, fast-frozen relations, with their train of ancient and venerable prejudices and opinions, are swept away, all new-formed ones become antiquated before they can ossify. All that is solid melts into air . . . ”³²

Building upon this famous metaphor, Zygmunt Bauman has described the development of this stage of modernity as one whereby, '[t]he melting of solids lead to the progressive untying of economy from its traditional political, ethical and cultural entanglements . . . [and thereby] . . . sedimented a new order, defined primarily in economic terms.'³³ The foremost casualty

³⁰ A view shared by Joseph Schumpeter, though for reasons aligned with Dicey rather than Marx, in that he believed that the modern corporation 'relentlessly narrows the scope of capitalist motivation': J. Schumpeter, *Capitalism, Socialism and Democracy* (1994, first published 1943) 156.

³¹ K. Marx, *Capital: A Critique of Political Economy* Vol. III (1981) chap.27, and p.569. This is not a proposition which would be recognised by economic liberals for whom the recognition of shares as a form of private property has significant governance implications: see, e.g. E. Sternberg, *Corporate Governance: Accountability in the Marketplace* (2nd ed., 2004). However, this view of shareholding would seem untenable, certainly in large publicly traded companies: see further, P. Ireland, 'Company Law and the Myth of Shareholder Ownership' (1999) 62 *M.L.R.* 32 and J. Parkinson, *Corporate Power and Responsibility* (1993) 32-41.

³² K. Marx and F. Engels, *The Communist Manifesto* (1998) 38. For an incisive account of the relationship between this text and modernity see M. Berman, *All that is Solid Melts Into Air: The Experience of Modernity* (1983) Part II.

³³ Z. Bauman, *Liquid Modernity* (2000) 4 (hereafter '*Liquid Modernity*'). The following account draws heavily upon Bauman's analysis in this work (especially the foreword and chapters 1, 2 and 4) and also upon the related essays in Z. Bauman, *The Individualized Society* (2001). Many of the issues raised are the

of this dissolution was the conception of being born into a place and status which essentially determined an individual's subsequent life-course through the interaction of a dense constellation of traditional cultural norms and customary practices. The emergence of a capitalist order thus generated new possibilities for individuals but the price was that it simultaneously, 'left the whole complex network of social relations unstuck - bare, unprotected, unarmed and exposed, impotent to resist the business-inspired rules of action and business shaped criteria of rationality, let alone to compete with them effectively.'³⁴ The economic logic of the process, however, initially resulted in the evolution of a stabilisation mechanism by which the freedom imposed upon individuals by the decline of tradition was directed through the construction of distinct class and gender roles that managed expectations and re-constrained action. In this manner these constructs provided a means by which the old solids could be replaced by new solids which could reasonably be expected to be durable in a world of radical and rapid transformation; such new solids thereby constituting a regime (to use Bauman's phraseology) of heavy capitalism or solid modernity.³⁵

An undoubted iconic symbol of heavy capitalism is to be found in the mass production system of the Fordist factory, as exemplified by Henry Ford's plant at Highland Park in San Francisco.³⁶ It embodies all the key features of solid modernity, viz: the utilization of large centralized organizations governed by instrumental rationality³⁷ and directed towards mass production utilising economies of scale requiring considerable investment in heavy plant with assembly lines populated by a whole host of labourers each undertaking a discrete largely de-skilled operation in a production process orientated around Taylorist principles.³⁸ The primary emphasis is thus upon permanence and structure and the result is a prison that neither capital nor labour is able to leave.³⁹

subject of considerable contention and the principal areas of disputation will be highlighted here or in the later discussion of post-Fordism.

³⁴ *Liquid Modernity*, p.4.

³⁵ Such change was very thorough-going; time itself was restructured to fit with the regularities industrial production required, see E. P. Thompson, 'Time, Work-Discipline and Industrial Capitalism' (1967) 38 *Past and Present* 56.

³⁶ For introductions to Fordism see P. Brown and H. Lauder, *Capitalism and Social Progress* (2001) chap. 3 and D. Harvey, *The Condition of Postmodernity* (1990) chap. 8. The acme of this line of thinking may perhaps be found in J. K. Galbraith's concept of the technostructure (with its intimate relationship to the large corporation) in *The New Industrial State* (2nd ed., 1991).

³⁷ Given the planning and co-ordination problems Fordism entailed it is unsurprising that bureaucracy is strongly associated with it. For a nuanced defence of the 'ethos of office' against both philosophical and managerialist critiques see P. Du Gay, *In Praise of Bureaucracy* (2000).

³⁸ F. W. Taylor, *Principles of Scientific Management* (1947, first published 1911). Taylor's primary aim was to increase productivity by reducing 'soldiering' (deliberate slacking); efficiency was maximized by detailed time and motion studies which removed discretion from employees in the execution of their work. Taylor himself maintained a rather utopian vision of his project as he believed the resultant gains would be so plentiful as to ensure a co-operative approach within the workplace and thus obviate the need for unions or collective bargaining: see *ibid.*, *Testimony Before the Special House Committee in 1912*.

³⁹ *Liquid Modernity*, p.145.

In this environment capital and workers were thus locked together in a mutual interdependency based upon the economics of mass production, whereby the concentration of capital within large corporations was to a large extent mirrored by the agglomeration of workers needed to operate the plant. Within this seemingly permanent framework of long-term relations it made rational sense for the workers to act together so as to improve their common future lot by combining so as to exercise a countervailing collective power to that of capital.⁴⁰ Further, the movement of people into cities to feed the industrial maw created a problem of order and thus the state also became increasingly enmeshed into the equation. In this light, the subsequent adoption of Keynesian inspired economic policy and the construction of the welfare state, as the principal elements of the post-war settlement, were by no means coincidental as the state could use these stabilising capacities to avoid or at least ameliorate the consequences of the periodic crises or cyclical downturns in the economy. Hence, a protean and rather loose corporatist approach emerged, albeit bounded within the clear parameters of a national economy and society, as a means of seeking to encourage both growth and stability.⁴¹

However, as the aforementioned quote from *The Communist Manifesto* clearly indicates, the dynamics of capitalism are relentless and despite the formidable institutional edifices assembled it proved impossible to maintain the Keynesian-Fordist regime. Instead, through the rise of an informational society⁴² and the concomitant processes of globalization national boundaries have been rendered evermore permeable and collective ties have been considerably weakened. In contrast to the regimented and predictable world of heavy capitalism instantaneous information flows and borderless competition have generated a paramount need for speed and flexibility: in this world of short-term horizons modernity has been rendered liquid and capitalism light.⁴³

For Bauman there appear to be two central characteristics of light capitalism: first, power is exercised foremost by mobility, and secondly, people are subjected to radical 'individualization'. As the combination of these ideas has profound consequences for the notion of collectivism here under examination it is proposed to analyse them in some detail. By 'individualization' Bauman refers to the fact that individuals living under present conditions have no given identities, (traditionally ascribed roles having been stripped away in the acid bath of modernity) and are, therefore,

⁴⁰ *ibid.*, p.33.

⁴¹ Clearly the United Kingdom has never come close to developing a truly corporatist system; the high-water mark was reached with the publication of the ill-fated Bullock Report: *Committee of Inquiry on Industrial Democracy* (1977; Cmnd. 6706). For a well contextualized account of company law in this period see T. Hadden, *Company Law and Capitalism* (2nd ed., 1977), especially Parts I and III.

⁴² The term refers to a society permeated by new information technologies which are themselves pivotal sources of productivity and power, see further M. Castells, *The Rise of the Network Society* (2nd ed., 2000) 21, n.31.

⁴³ Bauman respectively equates the period of solid modernity with heavy capitalism and liquid modernity with light capitalism, e.g. *Liquid Modernity*, p.145 and p.167.

forced to construct them for themselves.⁴⁴ Of course, being inherent to modernity, individualization was a feature of heavy modernity too, but crucially the surrogates of class and gender that served as frames for personal and collective identity orientation during the Fordist period are no longer straightforwardly available under the regime of liquid modernity.⁴⁵ The reasons for this are complex, but in essence it would seem that the premium currently placed on flexibility leads to high degrees of uncertainty, low levels of trust and a focus on short-term engagements. These factors, in turn, reduce the ability to have a secure and comprehensive, as opposed to provisional and partial, grasp of the present and thus render any strong teleological view of societal progression incoherent. Hence, in stark contradistinction to Dicey merely a century earlier, it is Bauman's view that, '[t]hrough the idea of improvement . . . through the legislative action of society as a whole has not been completely abandoned the emphasis (together with importantly the burden of responsibility) has shifted decisively towards the self-assertion of the individual.'⁴⁶

Further, the demise of collectivism is compounded by the onerous nature of the responsibility placed upon each individual and the unequal allocation of resources to perform the task. As Bauman notes:

“[t]he task of self-identification has sharply disruptive side effects. It becomes the focus of conflicts and triggers mutually incompatible drives. Since the task shared by all has to be performed by each under sharply different conditions, it divides human situations and prompts cut-throat competition rather than unifying a human condition inclined to generate co-operation and solidarity.”⁴⁷

However, even as the possibilities of collective action for labour seem to be receding, as it becomes ever more difficult for individuals' interests to cohere, it is also the case that labour is simultaneously losing out in the dimension of power. The reason for this is that labour has largely remained bound to particular local geographical territory whilst capital alone has broken free of the previously binding relationships that constituted the matrix of solid modernity and attained unprecedented mobility within the deregulated flows of the global marketplace.⁴⁸ As power in the context of liquid modernity lies in the ability to disengage even the threat of withdrawal based

⁴⁴ The concept is thus quite different to *laissez-faire* individualism; see also U. Beck and E. Beck-Gernsheim, *Individualization* (2002) preface xxi, and 202.

⁴⁵ For Beck class in second modernity is indeed reduced to a zombie category, *ibid.*, p. 206. Less radically, a recent historical study has discerned three basic models of class (labelled hierarchical, tripartite and dichotomous respectively) which intersect and operate simultaneously in the contemporary British consciousness but nevertheless concludes that whilst class continues to have significance this is not primarily as a mobilizing collective force: see further D. Cannadine, *Class in Britain* (2000).

⁴⁶ *Liquid Modernity*, p.29.

⁴⁷ *ibid.*, p.90, and see also p.35. Truly, to use Milan Kundera's phrase, an 'unbearable lightness of being'.

⁴⁸ This is certainly true of the global financial markets but is taken by many commentators to apply equally to transnational corporations, see *e.g.* G. Soros, *George Soros On Globalization* (2002) 1.

upon the real potential to do so creates a powerful bargaining chip for capital against both workers and state. Capital flight is thus one of the principal organising features of light capitalism and the resultant accommodation of the expanded market its accompanying *leitmotif*.⁴⁹

Further, even the requirements of greater flexibility demanded by the regime of light capitalism do not appear to have undercut the resilience of the large corporation by breaking the huge monoliths that populated the Fordist landscape into more nimble small and medium enterprises as some commentators had predicted.⁵⁰ Hence, the demise of collectivism has impacted differentially upon the Fordist trinity by re-configuring the bounds of the market and public space, very crudely: workers are individualized; the state loses many of its substantive directive capacities and increasingly undertakes an enabling role; and companies seem increasingly powerful as institutional repositories of collective resources⁵¹ subject principally to a market discipline in any event closely aligned with the rationalities and values underlying their own constitution.⁵²

Individualism Re-born: Thatcherism and the Neo-Liberal Heritage

In order to better understand the way in which companies are conceptualised within the territory of light capitalism it is also necessary to trace the development of the intellectual currents opposed to collectivism during the twentieth century and Dicey once again provides a good point for departure. At the outset of the twentieth century Dicey was already nostalgic for the days of Benthamite liberal individualism (which had been in its pomp some forty years before), even though he was self-consciously writing his analysis of the growth of the collectivism when that expansion was yet to reach its zenith - his preface to the second edition of *Law and Public Opinion in England* (published in 1914) indicates his awareness that the corporate entity, trades unions and the state itself were all still expanding both their size and capacities. Indeed, it was perhaps due to the relatively nascent development of collectivism and his acknowledgement that some social reform was necessary that Dicey's outlook as a committed liberal individualist was manifested only by a general wistful tone and occasional

⁴⁹ The reality of capital flight is a matter of considerable contention; Bauman adopts a hyperglobalist stance with little qualification: *Liquid Modernity*, pp.149-150. Commentators more sceptical of globalization view companies as less likely to seek extraterritoriality and nation states as retaining greater capacities: see, e.g. P. Hirst and G. Thompson, *Globalization in Question* (2nd ed., 1999) chap. 8.

⁵⁰ For a brief discussion see Castells, *op. cit.*, n.42, pp.167-8.

⁵¹ In the field of economic regulation, where substantial organizations (including the state and state agencies) bargain amongst themselves in 'regulatory space', this facet has made the large company both a very important and, given that by virtue of its functions it transcends the conventional public-private divide, also a unique player. See further, L. Hancher and M. Moran eds., *Capitalism, Culture, and Economic Regulation* (1989) chap. 10, and especially 272 and 274-5.

⁵² This is not to suggest that economic efficiency is necessarily the principal driver behind the development of the large company: for alternative political and organizational accounts see respectively, W. G. Roy, *Socializing Capital: The Rise of the Large Industrial Corporation in America* (1997) and C. Perrow, *Organizing America: Wealth Power and the Origins of Corporate Capitalism* (2002).

warnings as to the consequences of the collective principle if worked through to its logical conclusion.

However, by 1944, when Friedrich Hayek published *The Road to Serfdom*, the appeal to individualism had become much more passionate and urgent as collectivism was much more firmly entrenched and indeed the liberal influence of writers such as Dicey had waned to an extent that caused Hayek to lament that such figures once ‘. . . admired in the world at large as outstanding examples of the political wisdom of liberal England, are to the present generation largely obsolete Victorians.’⁵³ For Hayek, writing against a background of the totalitarian regimes of Nationalism Socialism in Germany and Stalinist Marxist Leninism in the USSR, this generational change in outlook was most regrettable, as it was the dangers of collectivism (particularly the threat to individual freedom highlighted by the like of Dicey) which were now all too apparent.

Although the precise definition of the concept of collectivism was problematic for Hayek, as was the case for Dicey, it is clearly focused upon socialism and the centralised control of the economy by the state.⁵⁴ In Hayek’s view at a technical level planning on this basis was doomed to inefficiency on epistemological grounds as the state was incapable of collating the information required, and at a moral level such co-ordination was wrong as it was imposed upon individuals regardless of their own values and desires.⁵⁵ Ultimately there was thus a stark moral choice to be made in selecting a method of economic co-ordination between the paths of freedom and coercion: liberalism gave priority to the former and collectivism inevitably resulted in the latter. For Hayek the key policy prescription to preserve individual liberty was thus to maximise use of the spontaneous forces of society and to resort as little as possible to coercion. However, although economic co-ordination was to be provided primarily by the spontaneous workings of the free market, Hayek repeatedly rejected a dogmatic *laissez-faire* stance and indicated that the state does have a role (albeit relatively minimal) particularly in maintaining an effective legal system and preserving competition.⁵⁶

The political facts then underlying world events meant that Hayek’s principal argument was directed to the relation between the state and individual and, whilst this axis brought the market to the fore, it did not have great deal to say about the significance of the corporate entity. Further, the collective

⁵³ F. A. Hayek, *The Road to Serfdom* (1997, first published 1944) 136. It is not possible in this context to examine the rich vein of work Hayek subsequently produced, notably *The Constitution of Liberty* (1960) and *Law, Legislation and Liberty* (1993). For a most insightful survey and critique of Hayek’s oeuvre see J. Gray, *Hayek on Liberty* (3rd ed., 1998).

⁵⁴ The problem is of quite a different order to that posed by Dicey; see Hayek, *The Road to Serfdom*, pp.24-26.

⁵⁵ The two strands are intimately related given Hayek’s belief that knowledge is tacit and, therefore, local.

⁵⁶ E.g. *ibid.*, pp.27-29 and p.60. Whilst here both of these functions are seemingly conceived in part as self-consciously creative acts by the state, in later writings Hayek seems to suggest an application of the principle of spontaneous order to the constitution of the institutional fabric itself: for a brief commentary see Gray, *op. cit.*, n.53, p.70.

effort and technological progress evidenced during the war stimulated an expansive and optimistic view of state action and Hayek's ideas thus fell on largely stony ground.⁵⁷ In addition, Hayek's views in *The Road to Serfdom* were perhaps somewhat negatively expressed.⁵⁸ It fell to Milton Friedman, some twenty years later, to place the populist liberal economic argument concerning competitive capitalism in a more celebratory mode, with the publication of *Capitalism and Freedom* in 1962.⁵⁹ Friedman explicitly aligned himself to the tradition of Dicey and Hayek and consequently also sought to restrict the ambit of the state and to maximise the freedom of the individual through the mechanism of market exchange. In Friedman's view contracts determined in a free market provided the optimal level of coordination without resort to coercion as it is the case that, '... both parties to an economic transaction benefit from it, *provided the transaction is bilaterally voluntary and informed.*'⁶⁰ The role of the state was, in similar vein to Hayek, to establish and enforce the legal framework, to act paternalistically in the rare case of mental incapacity and to intervene in cases of monopoly.

The discussion of monopoly power is central to Friedman's exposition of competitive capitalism and involves an examination of the Fordist triad of corporation, labour and state. The reason that monopoly is so important within Friedman's framework is due to the fact that it removes choice with the consequent dual effects of preventing true free exchange and of rendering the monopolist's power visible, thereby leading to calls for it to be exercised in a socially responsible manner. As is well known, Friedman strongly rejects any suggestion that the agents of business (or labour) have in general any duty to act responsibly or in any way beyond that served by the pursuit of the shareholders (for unions read members) self-interest. Fortunately his survey of the extent of monopoly in the United States proved that there was little overall cause for concern in the business field and, whilst growing, the monopolistic effects of labour combination were similarly still fairly negligible: the major problem lay in government regulation through federal agencies of sectors such as agriculture and transport.⁶¹ Thus it would seem that Friedman's fundamental position, as verified by him on a general empirical level, can be accurately summarised as follows:

"The participant in a competitive market has no appreciable power to alter the terms of exchange; he is hardly visible as a separate entity; hence it is hard to argue that he has any 'social responsibility' except that which is shared by all citizens to obey the law of the land and to live according to his lights."⁶²

⁵⁷ The 1945 general election resulted in the Labour Party achieving a landslide victory and the Atlee government proceeded to lay the foundations of the modern welfare state.

⁵⁸ The title itself sets the oppositional tenor of the book and also establishes its dark and sombre tone.

⁵⁹ M. Friedman, *Capitalism and Freedom* (1982, first published 1962).

⁶⁰ *ibid.*, p.13 (italics in original).

⁶¹ *ibid.*, chap. VIII.

⁶² *ibid.*, p.120: the parallel with Adam Smith's invisible hand is obvious, but note also Bauman's comment that, '[t]he favourite strategic principles of the powers-

Although this is not a point that Friedman needed to elaborate upon, it would seem from his discussion of enterprises, made in the immediate textual context of the above quoted proposition, that a company would constitute such a ‘virtually invisible’ market participant.⁶³ Further, this conclusion is reached despite the express consideration (admittedly made earlier and in a different context) of the application to the corporation of the conflict arising between the freedom to combine and the freedom to compete.

Dicey also made reference to this point and it is instructive to note the different ways in which these two liberal thinkers frame the issue and how the notion of collective power is much less visible and important in Friedman’s analysis. For Friedman the conflict is essentially one within the market domain and concerns the possibility of a monopoly arising either through restriction of the freedom to establish an enterprise or restriction of access to the market by established participants.⁶⁴ Whilst Dicey’s discussion of the characteristics of modern commerce is also concerned with monopoly,⁶⁵ he nevertheless situates the narrow problem at hand squarely within his conceptual opposition between collectivism and individualism. In this light the freedom to incorporate with limited liability that was finally consolidated in the Companies Act 1862 paradoxically both enhanced individual freedom and yet also simultaneously undercut it by transferring the management of business to corporate bodies, and thereby substituting ‘combined for individual action’.⁶⁶

The work of Hayek and Friedman thus provided a set of ideas which linked economic individualism with political freedom and which elevated the market into the key institutional mechanism for ‘organizing’ society. By way of correlation, the state’s role, together with that of the political realm generally, was to be consciously reduced by the adoption of ‘. . . a presumption or prejudice in favour of individual liberty, that is, of *laissez faire*.’⁶⁷ Within the market equity would be realised by formal rules of equality of opportunity (with no regard to equality of outcome)⁶⁸ and, in the absence of monopoly, market actors would be treated alike with no reference to their relative substantive capacities. As Friedman himself dryly noted, such ideas were somewhat out of the mainstream given the predominant

that-be are nowadays escape, avoidance and disengagement, and their ideal condition is invisibility’: *Liquid Modernity*, p.40.

⁶³ This ambiguity is nevertheless unfortunate given the undoubted significance of the company within the system of competitive capitalism. Contrary to the orthodox legal understanding Friedman appears to view companies as direct instruments of their shareholders with no effective separate personality (*op. cit.*, n.59, p.135) and private companies are on this basis directly equated with individuals in the market (*ibid.*, p.14). The position of publicly traded companies is hard to glean but presumably they too must be regarded as normal market participants unless they have attained a monopoly position.

⁶⁴ *ibid.*, p.26.

⁶⁵ Like Friedman who in most cases of technical monopoly would reluctantly opt for the least worse evil of private ownership (*ibid.*, p.28) Dicey would seem reluctant to endorse state ownership as preferable to joint stock ownership in such circumstances, see Dicey, *op. cit.*, n.18, pp.245-248.

⁶⁶ *ibid.*, p.246, n.2.

⁶⁷ Friedman, *op. cit.*, n.59, p.201.

⁶⁸ *ibid.*, p.195.

Keynesian mode of thinking at height of Fordist collectivism and thus received little attention until experience changed the climate of thinking.⁶⁹

The events that led to such a change are relatively easy to list (principally the oil shocks, high inflation, and industrial conflict of the 1970s and the nascent technological and cultural changes of that decade, which would lead to an intensification of competition in the global market, together with an increasing emphasis on consumption as opposed to production) but notoriously difficult to analyse, particularly as to causality, given the number of inter-related processes.⁷⁰ Further, whilst many companies have employed lean just-in-time production and sought niche markets, it is clear that elements of Fordist mass production still remain, even if production is more automated and companies more downsized and delayered. Nonetheless, whether the new environment is labelled as a regime of flexible accumulation,⁷¹ disorganized capitalism,⁷² light capitalism or post-Fordism,⁷³ one common feature is an acknowledgement of the decline of the economic nationalism that was at the core of the Keynesian-Fordist state and by implication of the unmediated directive capacities of the state itself.

The times were thus propitious for the Thatcherite ‘project’, which celebrated the individual and the market, and was explicitly both anti-statist and anti-collectivist (at least in relation to trades unions) in outlook. Whilst there is considerable debate as to extent that Thatcherism was conceived from the outset as coherent political project and to which it completely reflected the emerging environment as opposed to seizing upon certain elements of it, there is no doubt that it resulted in the ‘establishment of a neo-liberal economic and political agenda as the dominant paradigm to characterise British politics.’⁷⁴ Thatcherism drew heavily upon both Hayek and Friedman in order to provide a theoretical underpinning for the retreat of the corporatist state and the ascension of the market. As is well known, this required the nurturing of an enterprise culture that extended beyond the formal economy and required the construction of rugged individuals motivated by self-interest and capable of exploiting opportunities in a dynamic economy. The enterprise culture and the market-individualism it demanded was thus ‘. . . at the very heart of the Government’s moral crusade.’⁷⁵

⁶⁹ *ibid.*, 1982 Preface.

⁷⁰ See A. Amin ed., *post-Fordism- A Reader* (1994). In his introductory chapter Amin indicates the dispute as to the very existence of the transition and discusses three of the leading post-Fordist models, *viz*: the regulation school, the neo-Schumpeterian school and the flexible specialization school.

⁷¹ Harvey, *op. cit.*, n.36, Part II.

⁷² S. Lash and J. Urry, *The End of Organized Capitalism* (1987).

⁷³ For accessible overviews see, Brown and Lauder, *op. cit.*, n.36, Part II, and A. Stewart, *Theories of Power and Domination* (2001) chaps. 3 and 4.

⁷⁴ For a subtle analysis, see R. Heffernan, *New Labour and Thatcherism* (2001) chaps. 3 and 4 (quotation p.60).

⁷⁵ See P. Heelas and P. Morris eds., *The Values of the Enterprise Culture* (1992) (quotation p.1) for an excellent set of critiques. Chapter 2 of this work also contains a short discussion of the enterprise culture by one of its chief progenitors, Lord Young.

The pursuit of this objective led to some deep paradoxes, for, as discussed above, the essential dynamism of the capitalist process is such as to place societal authority, coherence and stability in constant question due to the new freedoms it undoubtedly generates. Hence, a strong state is required in order to make appropriate regulatory interventions. By way of illustration, markets in complex societies require to be structured in such a way as to ensure their visible probity, so as to maintain confidence in their operation. An excellent example of this is given by Bruce Carruthers and Terence Halliday in their study of the regulation of the insolvency market by the Thatcher government. The mode of governance adopted essentially involved licensing a monopoly position to a newly created profession of insolvency practitioners. Whilst this would on general policy grounds have been a complete anathema, it was here a necessary evil, allowing an indirect extension of the state's power of surveillance into a private market through the reporting obligations placed upon the profession. Contrary to neo-liberal theory, in the cold light of reality markets do not therefore spontaneously emerge but require to be actively constituted for '[t]he legitimation of markets, and the provision of normative foundations, is substantially a political problem.'⁷⁶

In addition, the intensification and greater permeation of market operations within society and the expansion of market processes and attitudes that an enterprise culture entailed generated a deeper paradox, for as Stefan Breuer has noted, '[m]arket societalization means an increase in interdependency and the atomization of the social, the increasing density and the negation of all ties - asocial sociability.'⁷⁷ In this climate the outcome of competitive market processes was thus more likely to produce an aggressive market individualism focused around calculative behaviour, unrestrained egoism, ambition and excessive consumption than the more restrained Victorian virtues of prudence, thrift and respectability that many Thatcherites espoused. A further facet of the 'asocial sociability' underlying market individualism derives from the principal theoretical basis the Thatcherites adopted, for '[n]eoliberal economics rests upon an image of the autarkic human self. It assumes that individuals alone can master the whole of their lives, that they derive and renew their capacity for action within themselves.'⁷⁸ Accordingly, for Thatcherites the avowed belief in the justice of market outcomes as the result of a non-coercive process thus enabled considerable moral responsibility to be placed upon the individual for any setbacks: failure was due to a lack of personal effort or enterprise. Such an all encompassing view of market relations led to a further paradox, for (being morally neutral as to outcome) the market is itself incapable of generating

⁷⁶ T. C. Halliday and B.G. Carruthers, 'The Moral Regulation of Markets: Professions, Privatization and the English Insolvency Act 1986' (1996) 21 *Accounting, Organizations and Society* 371, 373. The ramifications of this are far reaching for as Campbell (who strongly endorses the value of market arrangements) states, '[t]he private, as a category in itself, does not exist, for the private is a public construction, the limits of which are then a matter of public debate': D. Campbell, 'The Hybrid Contract and the Merging of the Public and Private Law of the Allocation of Economic Goods' in D. Campbell and N.D. Lewis eds., *Promoting Participation: Law or Politics?* (1999) 68.

⁷⁷ S. Breuer, 'The Denouements of Civilisation: Elias and Modernity' (1991) 128 *International Social Science Journal* 401, 407.

⁷⁸ Beck and Beck-Gernsheim, *op. cit.*, n.44, preface xii.

the underpinning values and norms upon which its very operation depends. Further, the social cohesion necessary to undertake this task is particularly hard to build if individuals are regarded as entirely self-sufficient (which in any event is obviously empirically problematic) as ‘trust is far more likely to flourish in a society whose members openly acknowledge their interdependencies and vulnerabilities and build their ethic of fairness around this realization.’⁷⁹

In the light of the above, and the central paradox that the Thatcherites embraced a radical view of market relations whilst simultaneously seeking to maintain traditional values and gender roles, it is unsurprising that there was a resort to ‘authoritarian populism’⁸⁰ in order to maintain social stability. However, whilst the state thus needed to be strong, it was also undoubtedly re-configured to a more enabling role, and corporatism and trades union collectivism were effectively swept away in accordance with Hayek’s precepts that,

“... the real exploiters in our present society are not egoistic capitalists or entrepreneurs, and in fact not separate individuals, but organizations which derive their power from the moral support of collective action... More real injustice is probably done in the name of group loyalty than from any selfish motives.”⁸¹

The ground was thus cleared for the market, and most importantly the values embodied in the market individual, to hold sway: for as Lady Thatcher herself made clear whilst, ‘[e]conomics are the method: the object is to change the soul.’⁸² At the same time the collective power of the corporation was rendered invisible as the company was seemingly absorbed as just one other actor into the anonymity of the market process.⁸³

‘New Times’ and New Labour

The demise of collectivism, as exemplified by organised labour and the Keynesian Fordist state, together with the radical re-vivification of the individual as a *homo economicus* in the Thatcherite neo-liberal mould has placed a great deal of expectation upon the market as an institutional tool and a corresponding pressure on individuals to adopt an entrepreneurial self, for as David Jenkins has put it ‘[a] useful *modus operandi* has been overpromoted into being an inevitable *modus vivendi*.’⁸⁴ However, New

⁷⁹ L.E. Mitchell, *Stacked Deck: A Story of Selfishness in America* (1998), especially chap. 10 (quotation p. 208). See also M. A. Fineman, *The Autonomy Myth* (2004) urging collective responsibility in the family sphere as an acknowledgement of universal dependency.

⁸⁰ For general critiques see: S. Hall and M. Jacques eds., *The Politics of Thatcherism* (1983) (quotation p.10); Giddens, *Beyond Left and Right* (1994) chap. 1; J. Gray, *False Dawn* (1999) 24-28; and D. Marquand, *Decline of the Public* (2004) chap. 4.

⁸¹ Hayek, *Law, Legislation and Liberty* (1993) Vol. 3, 96.

⁸² *The Sunday Times*, 7 May 1988.

⁸³ Hayek’s later views are most instructive here, for although he identifies organizational power as both important and undesirable (due to its essentially coercive nature), he is most reluctant to apply his analysis to the company as opposed to the trade union, see *op. cit.*, n.81, chap.15.

⁸⁴ D. Jenkins, *Market Whys and Human Wherefores* (2000) 114.

Labour has not relieved this pressure on individuals nor refuted the need for markets, as both are deemed inescapable if Britain is to be competitive in the dynamic, globalized world economy: in Gordon Brown's words ' . . . our mission must be relentless: to strengthen markets, to maximise efficiency.'⁸⁵ Hence, traditional Labour demand management policies have been discarded⁸⁶ in favour of supply-side interventions (which in Brown's view actively constitute growth according to neo-classical endogenous growth theory) in order to support a culture of enterprise and the structural flexibility required within a modern knowledge economy. The role of the state in the economy has thus been re-modulated, for, as Alan Finlayson has noted, whilst ' . . . dogma driven neo-liberals sought to liberate the market from the state, New Labour seeks to deploy that state in the name of the market because it sees that the nature of the market has changed.'⁸⁷ In this sense New Labour has both adopted and adapted the inherited Thatcherite terrain.⁸⁸

However, in contrast to Thatcherism, New Labour's vision of the world is not primarily ideological for,

“ . . . New Labour does not formulate political ideas on the basis of a substantial moral claim about the nature of society and the distribution of its resources. It does so on the basis of a 'sociological' claim about the novel condition of contemporary society; a belief that the world has been transformed, while our political ideas have not kept up pace . . . [Thus] [p]olicy is legitimated not by ethical principles but by the truth of certain social facts . . . ”⁸⁹

Finlayson identifies that two key influences on New Labour in determining the relevant social facts are the analysis of Thatcherism and post-Fordism given at the end of the 1980s in the 'New Times'⁹⁰ and the work of Anthony Giddens. The New Times analysis by leading Marxists sought to develop ideas for the left in response to 'epochal change' and roughly comprised two broad strands of thinking, one organisational and the other ideological, but

⁸⁵ Speech to the Social Market Foundation (3 February 2003). Whilst this speech contains a far reaching analysis of the role of markets, including the views of critics on their moral limitations, the fundamental message is accurately conveyed by the quotation. For a good summary of the present New Labour stance see the election speeches of Blair (Canary Wharf, 14 April 2005) and Brown (London Business School, 27 April 2005). For a comprehensive account of the New Labour project see S. Driver and L. Martell, *New Labour: Politics after Thatcherism* (1998) and *Blair's Britain* (2002).

⁸⁶ Such policies are perceived as infeasible in any event due to the electoral implications of taxation. A point presciently noted by Dicey as being likely to arrest collectivist growth: *op. cit.*, n.18, p.302, n.1.

⁸⁷ A. Finlayson, 'New Labour: the Culture of Government and the Government of Culture' in T. Bewes and J. Gilbert eds., *Cultural Capitalism: Politics after New Labour* (2000) 177, 186.

⁸⁸ The post-Thatcherite label should not disguise fundamental differences particularly over social welfare and cohesion; see Driver and Martell (1998), *op. cit.*, n.85, chaps.5 and 6.

⁸⁹ A. Finlayson, 'Third Way Theory' (1999) 70 (3) *The Political Quarterly* 271, 271.

⁹⁰ Hall and Jacques eds., *New Times: The Changing Face of Politics in the 1990's* (1989).

both based on the wider cultural understanding of political economy that emerged in the Thatcherite period.

The organizational thread⁹¹ has been the most influential and received an undoubted boost due to the exponential growth in information technology that has occurred since the early 1990s combined with an increasing awareness of the processes of globalization. Indeed, the need for flexibility and entrepreneurship in order to compete in a relentlessly dynamic global market driven by the knowledge economy and organized through open-ended networks has become a New Labour mantra.⁹² Whilst market individualism was thus embraced, albeit for practical as opposed to ideological purposes, New Labour was not blind to the problems of social cohesion that beset the Thatcherite years and the difficulties that rampant individualism had caused for the very sustenance of effective markets. Thus in order to socialise the market individualist, Blair first focused upon the concept of stakeholding,⁹³ which, as initially conceived, was a radical concept with application in the spheres of economy, society and politics.

An alternative approach to integrating the individual with society had already been offered by the ideological strand of the New Times which had analysed the rise in consumption and freedom associated with markets (in conjunction with the decline of collective solidarities) as opening up new ground for a radical politics of identity.⁹⁴ Such a politics would seek both to celebrate difference and diversity and to highlight their significance as hitherto neglected sites of conflict and power. In this respect there was a clear resonance with stakeholding, which, at the very least, operated by seeking to identify certain interests with particular individuals or groups and allowing for the pluralist representation of these interests. It would seem that this 'constitutionalized' conflictual element was the key conceptual problem with both ideas for New Labour, as the meta-picture of society that it was seeking *to present* was based on a fundamental identity of individual interest⁹⁵ framed by a globalization discourse of economic inevitability.⁹⁶ This approach is perhaps best illustrated by the New Labour account of social justice and wealth creation, which reconciles the traditional tension between

⁹¹ *ibid.*, see, e.g. the contributions by J. Urry and C. Leadbeater.

⁹² The exhilaration of light capitalism and the inevitable burden of individualization are well brought out in the titles to the principal works of the leading guru: see C. Leadbeater, *Living on Thin Air* (1999), and *Up the Down Escalator* (2002). A similar analysis has also been undertaken in the U.S.: see R. Reich, *The Work of Nations* (1993), and *The Future of Success* (2002).

⁹³ Launched in a series of speeches whilst still in opposition, see Blair, *New Britain: My Vision of a Young Country* (1996) Part III.

⁹⁴ See Hall, 'The Meaning of New Times' in Hall and Jacques *op. cit.*, n.90, p.116.

⁹⁵ This has led to a break with many post-Marxists for whom the essence of democratic politics lies in 'agonistic pluralism': see C. Mouffe, *The Democratic Paradox* (2000) chap.5, and Hall, 'The Great Moving Nowhere Show' *Marxism Today* (1998) 9.

⁹⁶ See M. Watson and C. Hay, 'The Discourse of Globalisation and the Logic of No Alternative: Rendering the Contingent Necessary in the Political Economy of New Labour' (2003) 31 *Policy & Politics* 289. For Watson and Hay this was a political choice but others on the left see the processes of globalization as genuinely constraining e.g. D. Coates, 'Capitalist Models and Social Democracy: the Case of New Labour' (2001) 3 *Brit. J. of Pol. & Int. Rel.* 284.

the two by opting for a model based on equality of opportunity, thereby consigning equality of outcome (with its dampening effect on enterprise and re-distributive focus) to history. This conception of social justice enables the state to portray much of its activity as a technocratic or managerialist response to structural changes affecting the whole of society,⁹⁷ whilst simultaneously devolving much of the responsibility for outcomes to individuals through their actions in the marketplace.

On this basis it is unsurprising that communitarianism has thus proved to be much more palatable than stakeholding as an organizing principle, for it allows issues to be addressed to an inclusive non-sectional audience, and ‘... offers Labour modernizers a political vocabulary which eschews market individualism, but not capitalism, and which imbues collective action but not class or the state.’⁹⁸ As such, the aspects of stakeholding that are too reminiscent of Old Labour, especially the corporatist and Keynesian elements in Will Hutton’s version,⁹⁹ are neatly avoided. However, the stakeholder concept was formulated very much with the large company in mind as a means of generating active space for the exercise of voice and the demonstration of accountability over corporate decisions.¹⁰⁰ Communitarianism by contrast, although also a broad concept which permits of many shades of meaning, has under New Labour tended to be of a conservative and prescriptive ilk.¹⁰¹ Further, whilst New Labour has been keen to stress the responsibilities of individuals (*e.g.* the principle of conditionality to welfare rights) within this framework,¹⁰² responsibilities in the corporate sector have been placed on a much more voluntarist footing within a context that has come to be dominated by the need for light regulation and low constraints on business actors.¹⁰³ In sum, the organizational strand of the New Times analysis has led New Labour to a world view dominated by global capital and in which community is primarily a resource for individuals to face technological imperatives rather than a collective public realm to generate countervailing values to those of market individualism.¹⁰⁴

Giddens and Post-Traditional Society

As the chief architect of the ‘Third Way’ there can be no doubt of Giddens’ influence upon New Labour.¹⁰⁵ That his politics spring from his

⁹⁷ See Finlayson, *op. cit.*, n.89, p.274.

⁹⁸ Driver and Martell, ‘New Labour: Culture and Economy’ in L. Ray and A. Sayer eds., *Culture and Economy after the Cultural Turn* (1999) 255.

⁹⁹ W. Hutton, *The State We’re In* (1995); *The State to Come* (1997); and *The Stakeholding Society* (1999).

¹⁰⁰ *ibid.* See also G. Kelly, D. Kelly and A. Gamble, eds., *Stakeholder Capitalism* (1997).

¹⁰¹ Driver and Martell, *op. cit.*, n.98.

¹⁰² See N. Rose, ‘Inventiveness in Politics’ (1999) 28 *Economy & Society* 467.

¹⁰³ See Blair’s speech to the IPPR (26 May 2005). For a discussion of how business interests led to preference accommodation (as opposed to preference shaping) by New Labour pursuant to the modified structural dependence thesis see Hay, *The Political Economy of New Labour* (1997) chap.5.

¹⁰⁴ See further, Marquand, *op. cit.*, n.80, chap.5.

¹⁰⁵ There are, however, clear differences between Blair and Giddens as the former takes a rather neo-liberal view of both globalization and individualism: see

understanding of radically new social facts, as Finlayson suggests, would seem equally incontestable:

“Revisionist social democracy starts from the reality of a structural transformation in the economy and society. It responds to intensifying global economic competition; the rise of a service and knowledge-based economy . . . and to the rise of a pervasive cultural individualism, itself linked to a decline in deference . . .”¹⁰⁶

However, neither stakeholding nor community figure significantly in Giddens’ account, which is determined principally by his investigation into the nature of the individual in high modernity.¹⁰⁷ The three key elements of his analysis are globalization, the emergence of a post-traditional order and the expansion of social reflexivity. The processes of globalization are important, not primarily in an economic sense, but due to the compression of time and space that they allow. Globalization thus entails continual disembedding, by which interaction is lifted out from the particularities of locales. Individuals are thus required to subject both their understanding and experience of a disembedded interaction to critical interrogation so as to re-assemble them reflexively, and thereby gain meaningful purchase on the social world. However, this very process also re-constitutes the interaction itself, and for Giddens post-traditional societies are thus irredeemably reflexive. Equally, such societies also operate at the edge of Enlightenment rationalism, as the far-reaching interventions so enabled (particularly in the scientific field) have generated manufactured risk, which, in turn, creates uncertainties not amenable to expert resolution.¹⁰⁸

The effect on the human self is quite profound. In traditional societies an individual’s space for action and ontological security was largely determined by received tradition and mediated authority (to a significant extent replaced by collective modes of assurance based upon Enlightenment rationality (*e.g.* bureaucracies) in early modernity). However, disembedding is hostile to received tradition, and thus in a post-traditional society it is in Giddens’ view necessary for each individual reflexively to create their own biography.¹⁰⁹ It is primarily in this frame that Giddens rejects both Keynesian economic management¹¹⁰ and neo-conservative communitarian attempts to ground the self.¹¹¹ Instead, the individual in high modernity needs to develop an autotelic self, which term:

“. . . refers to a person able to translate potential threats into rewarding challenges, someone who is able to turn entropy

further, Driver and Martell, ‘Left, Right and the Third Way’ in Giddens ed., (2001), *op. cit.*, n.2, pp.43-5.

¹⁰⁶ Giddens and Diamond eds., *op. cit.*, n.2, p.1. This position has been consistent from the outset: see Giddens (1998), *op. cit.*, n.2, p.26.

¹⁰⁷ The following account draws upon Giddens, *Modernity and Self-Identity* (1991) and *Beyond Left and Right* (1994) (hereafter ‘*Left and Right*’).

¹⁰⁸ Giddens (1991), *ibid.*, p.31.

¹⁰⁹ Class is thus no longer experienced simply as ‘collective fate’: *Left and Right*, pp.143-4.

¹¹⁰ It is hard to direct individuals in a reflexive world: *ibid.*, p.42.

¹¹¹ Such communities oppress individual autonomy: *ibid.*, pp.124-6.

into a consistent flow of experience . . . [and who] does not seek to neutralize risk . . . [which] is confronted as the active challenge which generates self-actualization.”¹¹²

Through the combination of this view of the self and the enabling role that is the primary avenue of action left to the state under a regime of reflexive modernity, Giddens also regards the broad field of politics as being radically changed. In early modernity the characteristic political mode was to be found in emancipatory politics, which sought to free individuals from tradition and mobilized collective power in order to achieve equal access to civil, political and social rights. By contrast, the focus in a period of reflexive modernity is perforce upon life politics, which is both generative and dialogical, and centres on questions of identity construction and lifestyle choice.¹¹³ Politics has, in this sense, thus moved beyond the conflictual collective left/right dichotomy and the structure of power in society has been accordingly re-configured.¹¹⁴ In essence, Giddens’ social democracy seeks to enable all members of society to participate in life politics, and this aim is reflected in a movement away from issues of substantive equality,¹¹⁵ to a concern with avoiding social exclusion, especially from the employment market.¹¹⁶

It is easy to see how a self conceived on such a basis potentially maps onto the market individualist required by neo-liberal enterprise culture. However, Giddens is clearly opposed to the reductive atomistic version of individualism propounded by neo-liberals, which he characterises as solely concerned with ‘. . . the self-seeking, profit-maximising behaviour of the marketplace.’¹¹⁷ Instead, for Giddens the core of individualism lies in embedded autonomy:

“[i]n a world of high reflexivity, an individual must achieve a certain degree of autonomy of action as a condition of being able to survive and forge a life; but autonomy is not the same as egotism and moreover implies reciprocity and interdependence. The issue of reconstructing social solidarities should not therefore be seen as one of protecting social

¹¹² *ibid.*, p.192. This notion of the self as both resolutely rational and infinitely plastic has been subject to critique: for a brief overview see A. Elliott, *Concepts of the Self* (2001) 36-45.

¹¹³ Giddens (1991), *op. cit.*, n.107, chap.7.

¹¹⁴ See, e.g. Giddens and Diamond eds., *op. cit.*, n.2, pp.106-8. It is precisely the fact that ‘[r]elations of power and their constitutive role in society are obliterated’ in Giddens’ thought, that in many critics eyes prevents Giddens from confronting the ‘. . . systemic problems of inequality and instability generated by capitalism’: see Mouffe, *op. cit.*, n.95, pp.110-12; and generally A. Callincos, *Against the Third Way* (2001).

¹¹⁵ See Giddens (2000), *op. cit.*, n.2, chap.4. For an argument that the rationale of life politics is too sharply distinguished from emancipatory issues and is liable to subversion by economic/market logic, see N. Mouzelis, ‘Reflexive Modernization and the Third Way: the Impasses of Giddens’ Social-Democratic Politics’ (2001) *The Sociological Review* 436.

¹¹⁶ Full employment is central to social justice: Giddens and Diamond eds., *op. cit.*, n.2, p.108.

¹¹⁷ *Left and Right*, p.13.

cohesion around the edges of an egoistic marketplace. It should be understood as one of reconciling autonomy and interdependence in the various spheres of social life, including the economic domain.”¹¹⁸

However, there would seem to be a fault line through Giddens’ thinking which rests on his ambivalent view of markets. For although markets bring freedom, dynamism, and innovation, the problems arising for social cohesion and solidarities are also clear.¹¹⁹ Ultimately, however, Giddens’ position is determined by the fact that there is no practical alternative.¹²⁰ It is thus made patent that the new egalitarianism proposed in his latest work pragmatically falls short of redistributive social justice, due to the need to accommodate economic imperatives.¹²¹ But such economic imperialism presents serious difficulties for the autotelic self and for a politics based upon values, for as Lawrence Mitchell has put it:

“The critical problem in the age of markets is that we have yet to develop appropriate norms, rules, and social structures to accompany it, social structures that are necessary to bring some stability and order . . . to the volatile and chaotic market environment in which we now live . . . The result is anxiety, resentment, fear, self-interest, and competition without regard to others. The result is the ethic of fear. And the ethic of fear threatens cooperation, compassion and fairness within our society.”¹²²

Mitchell concludes, rather ominously, that ‘[i]n the absence of consciously created social norms and social institutions to counter the destabilizing effects of the market, society itself is at risk of becoming nothing other than an economic arena.’¹²³

Such a danger is enhanced by the presence of the company as a dominant institution in society, with the sole *raison d’être* of being an efficient market actor. Indeed, Giddens’ early analysis of the new socio-political realities led him to muse that, ‘the globalizing of capitalist economic relations would seem on the face of things to leave large corporations in a dominant position within the economies of states and in the world economy as a whole’¹²⁴ However, having identified the potential of corporate collective power, Giddens concludes that this *prima facie* situation would in reality only present difficulties if the company obtained a monopoly. Whilst this conclusion has an unfortunate resonance with that of Friedman examined earlier, it is ventured that Giddens’ reasoning is more likely to be based upon

¹¹⁸ *ibid.*

¹¹⁹ *ibid.*, pp.10-11.

¹²⁰ In this respect, at least, Giddens frequently commends the truth of Francis Fukuyama’s, *End of History and the Last Man* (1992): see *e.g.* Hutton and Giddens eds., *op. cit.*, n.2, p.11.

¹²¹ Giddens and Diamond, *op. cit.*, n.2, pp.108-9.

¹²² Mitchell, ‘The Age of Aquarius or, How (I almost) Learned to Stop Worrying and Love Free Markets’ (2004) 88 *Minn. L. Rev.* 921, 949.

¹²³ *ibid.*, p.953. The problem is one that the Victorians also struggled with: see G. R. Searle, *Morality and the Market in Victorian Britain* (1998).

¹²⁴ *Left and Right*, p.89.

his general diffuse view of power and upon his specific understanding of the development of the flattened hierarchies characteristic of post-Fordist companies as a response to the need for greater reflexivity.¹²⁵

However, as his thinking has evolved, Giddens has become considerably less sanguine about the realities of corporate power. He has acknowledged that early progressive theory had been too accepting of business and that 'as a result of two decades of neoliberal dominance, the business world is threatening to erode its own mandate to operate.'¹²⁶ By way of a solution Giddens has started to explore the need for a civil economy and for the development of corporate citizenship and social responsibility.¹²⁷

Corporate Personality and Corporate Law Reform

It may be objected that much ink has already been spilt on the thorny topic of identifying the precise jurisprudential basis of corporate identity without generating any single agreed substantive model, and that any form of essentialist argument seeking to ground the nature of corporate personality is accordingly inappropriate and thus doomed to failure. Such was indeed the import of John Dewey's classic article in 1923, which amounted to a summation of the intellectual debate on the corporate form that had raged across the Atlantic at the beginning of the twentieth century, and in which Dewey's clear conclusion was that, 'for the purposes of law the conception of "person" is a legal conception; put roughly, "person" signifies what law makes it signify'.¹²⁸ However, Dewey's central argument is not fatal to the nature of the investigation envisaged here, which, instead, builds upon his wider observation that any theory of corporate personality is infused with materials 'logically extraneous' to the exercise of legal definition. Accordingly, being impressed with conflicts of both a political and an economic nature, corporate theory is inevitably plural.¹²⁹

For current purposes only a brief overview of this rich tapestry is possible, and it is proposed to concentrate upon three key domains: the nexus of contracts model; the doctrinal position; and stakeholder models. The nexus of contracts model of the company forms a significant element of the law and

¹²⁵ 'A post-bureaucratic organization can both harness social reflexivity and respond to situations of manufactured uncertainty much more effectively than a command system': *ibid.*, p.122.

¹²⁶ Giddens ed., (2003), *op. cit.*, n.2, p.11. On New Labour's need to address its 'overly cosy relationship with business' see P. Mandelson, *The Blair Revolution Revisited* (2002) p. xlvii.

¹²⁷ See Giddens and Diamond *op. cit.*, n.2, pp.112-3; Giddens ed., (2003), *op. cit.*, n.2, pp.6-12; and Giddens (2000) *op. cit.*, n.2, pp.142-153. An important contribution has also been made in developing the concept of an embedded market whereby the social values of communities are the principal regulators of economic life: see further, J. Kay, 'The Embedded Market' in Giddens ed., (2003), *ibid.*, pp.35-53; and *The Truth about Markets* (2003).

¹²⁸ J. Dewey, 'The Historic Background of Corporate Legal Personality' (1926) 35 *Yale L. J.* 655. See also H.L.A. Hart, *Essays in Jurisprudence and Philosophy* (1983) 21-48.

¹²⁹ See e.g. D. Millon, 'Theories of the Corporation' (1990) *Duke LJ* 201; and M. Hagar, 'Bodies Politic: the Progressive History of Organizational "Real Entity" Theory' (1989) 50 *Univ. Pitt. L. R.* 575.

economics scholarship that is currently the dominant mode of analysis in Anglo-American thinking on the subject¹³⁰ and has clearly been influential in the recent corporate law reform process.¹³¹ In essence, the company is conceived in contractual terms as a network of explicit and implicit bargains made between the participants in the enterprise. The analytical emphasis thereafter is placed upon dealing with the problems of the agency costs arising. In the model's purest form, the distinctive notion of the company as a separate legal person is thus rendered theoretically superfluous to the principal investigation.¹³² For present purposes the point of interest lies in this radical 'deconstruction' of the corporate entity, which arguably symbolically seeks to remove corporate collective power from the wider public political agenda, either as part of an explicit anti-regulatory neo-liberal approach, or more usually, so as to deal with the complex agency problems at the level of detailed microeconomics.¹³³

The extent to which the judiciary has been influenced by the nexus of contracts theory is not easy to ascertain. The lack of explicit reference to it within judgments is to be expected, as its abstract modelling is not easy to integrate with the traditional mode of reasoning employed in the common law. If the most that might therefore be found is an echo, then it is arguable that such a resonance may be detected in the leading modern case on corporate personality, *Meridian Global Funds Management Asia Ltd. v Securities Commission*.¹³⁴ The details of the case are not important, save that it concerned the need to attribute knowledge to a company. In his opinion on behalf of the Privy Council, Lord Hoffmann sought to move away from organic metaphors of the company¹³⁵ and move towards a set of primary and general rules of attribution drawn from the company's constitution and the general law respectively. In so doing Lord Hoffmann expressed the following view of corporate personality:

¹³⁰ See e.g. F. Easterbrook and D. Fischel, *The Economic Structure of Corporate Law* (1991); and B.R. Cheffins, *Company Law: Theory, Structure and Operation* (1997). For critiques see e.g. Ireland, 'Property and Contract in Contemporary Corporate Theory' (2003) 23 L.S. 453; and W.W. Bratton, 'The Economic Structure of the Post-Contractual Corporation' (1992) 87 *Northwestern U. L. R.* 180.

¹³¹ See especially, *Company Directors: Regulating Conflicts of Interests and Formulating a Statement of Duties* (1999: Law. Com. No. 261; Scot. Law. Com. No. 173) Part II.

¹³² See the points noted in Cheffins, *op. cit.*, n.130, p.32.

¹³³ For recent valuable contributions to this analysis see: *ibid.*, and R. Kraakman *et al.*, *The Anatomy of Corporate Law* (2004). It is notable that in order to contextualize their agency analysis Hansmann and Kraakman have to explain corporate legal personality and limited liability by way of a property-based argument (by what they term affirmative and defensive asset partitioning respectively) and, as Ireland notes, that this move breaks open any model based on the essential 'private, contractual nature of corporations' to allow a more radical governance regime based on their social power: see Ireland, *op. cit.*, n.130, p.509.

¹³⁴ [1995] 2 A.C. 500 (P.C.).

¹³⁵ See e.g. the judgment of Denning LJ in *H.L. Bolton Engineering Ltd v T.J. Graham & Sons Ltd* [1957] 1 Q.B. 159, 172 (C.A.).

“. . . reference to a company ‘as such’ might suggest that there is something out there called the company of which one can meaningfully say that it can or cannot do something. There is in fact no such thing as the company as such . . . only the applicable rules [of attribution].”¹³⁶

On the particular facts of the case the relevant knowledge was attributed to the company by way of a special rule of attribution determined by a purposive construction of the relevant legislation. There are two paradoxes here: first, the notion of attribution requires an attributee; and secondly, in denying the existence of the company ‘as such’ Lord Hoffmann was actually extending the scope of corporate liability.¹³⁷

By way of contrast, stakeholding models of the company are generally premised upon a strong notion of the corporation as the seat of significant power over the various constituencies with an interest in its operations.¹³⁸ The theoretical focus is thus to ensure that the corporate form is analysed within the full social, economic and political context in which it functions, so as to ensure that appropriate governance mechanisms are in place. The primary responses to this enquiry have been to seek to place directors under a fiduciary duty to each stakeholder, to encourage formal employee participation in corporate decision-making, and to foster corporate social responsibility.¹³⁹ Stakeholding thus has distinctly corporatist overtones and, as such, has proved politically unattractive to New Labour.¹⁴⁰ In addition, these overtones are amplified by the ongoing debate between the individualist and corporatist modes of capitalism¹⁴¹ within the European Union,¹⁴² and the more general contemporary focus on corporate governance regimes as a site of national comparative economic advantage.¹⁴³ As might

¹³⁶ [1995] 2 A.C. 506H-507A.

¹³⁷ For commentaries on the decision see: N. Lacey, ‘Philosophical Foundations of the Common Law: Social not Metaphysical’ in J. Horder ed., *Oxford Essays in Jurisprudence* (2000) 17; and G. R. Sullivan, ‘The Attribution of Culpability to Limited Companies’ [1996] C.L.J. 515.

¹³⁸ See e.g. Parkinson, *op. cit.*, n.31; and Mitchell ed., *Progressive Corporate Law* (1995). For a critique, from a person sympathetic to both Third Way politics and a thoroughly contextualized view of the corporation as a social actor, see S. Wheeler, *Corporations and the Third Way* (2002) 29-33.

¹³⁹ See the influential pieces by M. Stokes, ‘Company Law and Legal Theory’ in W. Twining ed., *Legal Theory and Common Law* (1986) 155; and Lord Wedderburn, ‘The Social Responsibility of Companies’ (1985) 15 *Melbourne U. L. R.* 4.

¹⁴⁰ Despite early dalliances as noted above in the main text: see also J. Plender, *A Stake in the Future* (1997); and Driver and Martell (1998), *op. cit.*, n.85, pp.51-60.

¹⁴¹ For an early account see M. Albert, *Capitalism Against Capitalism* (1993).

¹⁴² It is a clear undercurrent in the rejuvenated Lisbon Agenda, where the social model seems to become less prominent in each of the following documents: *Facing the Challenge: The Lisbon Strategy for Growth and Employment* Report from the High Level Group chaired by Wim Kok (November 2004); Communication of the Commission to the European Council COM(2005) 24 (2 February 2005); and Brussels European Council Presidency Conclusions 7619/05 (23 March 2005).

¹⁴³ For a comprehensive overview of the debate see: K. J. Hopt *et al.*, eds., *Comparative Corporate Governance* (1998); J. A. McCahery *et al.*, eds.,

be imagined from the previous discussion the individualist model is favoured by the United Kingdom, as it is seen as more flexible and dynamic in a world of change.¹⁴⁴ The European social model, by contrast, although encouraging long-term investments, is perceived as too cumbersome, and as overly insulating various interlocking elites from wider accountability and market pressures.¹⁴⁵

All things considered it is thus hardly surprising that the government has not proceeded with stakeholder models in its proposals for company law reform. In fairness, the government is largely following the recommendations of the Independent Company Law Review that it instigated in 1998 with the object of ensuring that the United Kingdom had an appropriate regulatory framework to maintain a competitive corporate regime in the globalized economy of the new millennium.¹⁴⁶ The Review did in fact consider the basis of company law by reviewing the content of the company directors' duties; and generated two models: the enlightened shareholder model (which was generally agreed to represent the existing law that companies should be run with a priority to shareholders interests) and the pluralist model.¹⁴⁷ The latter was in effect stakeholding under a different guise and was based not only on the benefits of fostering co-operative long-term relationships but also on economic analysis. The underlying economic reasoning was that shareholders were not the only residual risk bearers in the company as other stakeholders also made firm specific investments and were unable to protect themselves by writing complete contracts. In order to maximise the company's overall wealth creation it was thus necessary to optimise each stakeholders' firm specific investment by allocating governance rights beyond simply the shareholders.¹⁴⁸

Whilst the pluralist model was rejected (principally due to the accountability and enforcement problems arising from a multi-fiduciary duty) its proponents could claim that it left its imprint upon the inclusive shareholder

Corporate Governance Regimes (2002); and J. N. Gordon and M. J. Roe eds., *Convergence and Persistence in Corporate Governance* (2004).

¹⁴⁴ Blair's speech to the European Parliament (23 June 2005) on the revitalized Lisbon Agenda attempts to mitigate the individualism but is clear that the social model requires modernization.

¹⁴⁵ See Giddens (2000), *op. cit.*, n.2, pp.151-3. Giddens' disagreement with Hutton on these issues is best brought out in their conversation reported in Hutton and Giddens eds., *op. cit.*, n.2, pp.1-51. Whilst some political scientists believe that the social model continues to offer comparative institutional advantage others are more sceptical about its continuing ability to generate outcomes for labour that are superior to individualist models, see respectively: P.A. Hall and D. Soskice eds., *Varieties of Capitalism* (2001) 1-68; and Coates, *Models of Capitalism* (2000) 259-60.

¹⁴⁶ The White Papers are set out in n.1. For a detailed commentary on the review process by its director, see J. Rickford, 'A History of the Company Law Review' in J. de Lacy ed., *The Reform of United Kingdom Company Law* (2002) 3.

¹⁴⁷ For details see *ibid.*, and G. Wilson, 'Business, State, and Community: 'Responsible Risk Takers'', New Labour and the Governance of Corporate Business' (2000) 27 *J. of Law & Soc* 151.

¹⁴⁸ See G. Kelly and J. Parkinson 'The Conceptual Foundations of the Company: a Pluralist Approach' in Parkinson *et al.*, eds., *The Political Economy of the Company* (2000) 113.

duty that was the outcome of the Review, and that has been adopted, in slightly modified form, by the government.¹⁴⁹ Pursuant to the proposed new statutory duty company directors must act in the manner they ‘consider, in good faith, would be most likely to promote the success of the company for the benefit of its members as a whole’.¹⁵⁰ In determining this they must, so far as reasonably practicable, take account of likely short and long term effects of the decision on the community and environment together with the interests of employees, customers, suppliers and the company’s own need to maintain a reputation for high standards of business conduct. As the primary duty is to enhance shareholder benefit and remains owed to the company, this is clearly not a stakeholder approach. However, although there may be some devil in the detail, it does, symbolically at least, open the company up to the outside environment and thus clearly renders it a visible legal actor (both to external and internal participants) therein. When combined with the regular disclosure of information by quoted companies under the new Operating and Financial Review, there is thus some cause for hope that corporate power will be exercised in a context more open to wider cultural values, and that reputational concerns in the market might be harnessed to ratchet-up corporate behaviour.¹⁵¹

However, for Sally Wheeler, the fact that the Review frames corporate relations ‘in a context of increasing competitiveness displays a worrying emphasis on the importance of the market as an institution divorced from wider concerns such as inclusion and shared advancement.’¹⁵² An ethical framework that fosters the latter values is vital if the company is to re-legitimate itself within society; and is central to Third Way philosophy, which seeks to re-unite the individual and the collective by way of voluntary action.¹⁵³ Wheeler, therefore, advocates that corporations should be treated seriously as individual actors and encouraged to participate in the *polis*, so as to realise themselves as virtuous citizens. Such corporate intervention (running in tandem with profit-making activities) would be actioned via partnership with Regional Development Agencies, assessed by way of need audits, and implemented through care plans.¹⁵⁴ The consequence would be

¹⁴⁹ Space does not permit a detailed examination of the primary sources, see further: Rickford, *op. cit.*, n.146, pp.15, 22 and 36-7; and Parkinson, ‘The Inclusive Company’ in de Lacy ed., *op. cit.*, n.146, p.43, especially pp.44-52. The proposed statement of directors’ duties is set out in the Corporate Law Reform Bill, clauses B1-8 (note that the detailed wording is significantly different to the previous draft Companies Bill in *Modernising Company Law* (July 2002; Cm 5553-II)).

¹⁵⁰ Corporate Law Reform Bill, clause B 3(1).

¹⁵¹ See the excellent discussion in Parkinson, ‘Disclosure and Corporate Social and Environmental Performance: Competitiveness and Enterprise in a Broader Social Frame’ (2003) 3 *J. Corp. Law Studies* 3.

¹⁵² Wheeler, *op. cit.*, n.138, p.51.

¹⁵³ Wheeler founds corporate ethical behaviour in Aristotelian virtue ethics and utilizes the work of Alasdair MacIntyre to generate a liberal localized conception of community.

¹⁵⁴ It is interesting to speculate whether situating the corporate form in this conceptual framework would ameliorate the massive disparity in general health that has been shown to beset societies in which there are large inequalities: see R. Wilkinson, ‘Social Corrosion, Inequity and Health’ in Giddens and Diamond eds., *op. cit.* n. 2, 183, and *The Impact of Inequality: How to Make Sick Societies*

that corporations would both practically apply the kind of values that New Labour seeks to espouse as a basis for progressive politics and see ‘themselves as contributing to the accumulated good of all other individuals and not as in some way hosted by a wider society and yet separate from it.’¹⁵⁵

Conclusion

It has been noted both, that the Labour Party has demonstrated surprisingly little interest in the underlying principles of company law during its history,¹⁵⁶ and that ‘[c]ompany law has been seen as an apolitical, technical subject concerned solely with the rules of private law’¹⁵⁷ However, for a progressive social democratic politics to prevail, and for the values and rationality of the market individualist not to successfully colonise other spheres of society, it is vital to recognise the power of corporate agency and for the company to be brought within dialogical engagement with other societal groups.¹⁵⁸ To this end the proposed company law reforms offer a chink of light in that company directors will be under a duty to at least identify the factors relevant to a corporate decision and the increased disclosure provisions will enhance the visibility of corporate activity. Nevertheless, whilst of some symbolic value, a significant limitation of the reforms may be said to lie in the fact that the enforcement of duties remains principally with the company itself. Taken as a whole the effect the reform package is thus that the corporation remains subject to discipline contoured principally by market rationalities.¹⁵⁹ Given New Labour’s tendency to adopt market solutions this outcome is doubtless not too surprising, but it is unfortunate that Giddens has not developed his notion of a civil economy with more urgency in the light of his increasing disquiet with aspects of corporate capitalism. It may be that Wheeler’s radical vision of the corporate entity provides the type of utopian thinking required, for otherwise the combination of largely unacknowledged corporate collective power and market individualism might yet result in a society such as that depicted by Kurkov’s novels on the anarchic capitalism of post-Soviet Ukraine: ‘It was all money, money, money today. Ideology was out.’¹⁶⁰

Healthier (2005). For a telling analysis of the inequalities arising from the shareholder-orientated model of the corporation, see Ireland, ‘Shareholder Primacy and the Distribution of Wealth’ (2005) 68 M.L.R. 49.

¹⁵⁵ Wheeler, *op. cit.*, n.138, p.80.

¹⁵⁶ See B. Clift, A. Gamble and M. Harris, ‘The Labour Party and the Company’ in Parkinson *et al.*, eds., *op. cit.*, n.148, p.51.

¹⁵⁷ C. Graham, ‘Regulating the Company’ in Hancher & Moran eds., *op. cit.*, n.51, p.199.

¹⁵⁸ Such dialogue has a double-edged quality, as it may also extend corporate influence: see Parkinson, ‘Participation and Corporate Governance’ in Campbell and Lewis eds., *op. cit.*, n.76, p.319, especially pp.344-7.

¹⁵⁹ The inadequacy of which is well demonstrated by Ireland, ‘Recontractualising the Corporation: Implicit Contract as Ideology’ in D. Campbell *et al.*, eds., *Implicit Dimensions of Contract* (2003) 255, especially 287-8. See also the recent polemical dissection of corporate power in the United States by Galbraith in *The Economics of Innocent Fraud* (2004).

¹⁶⁰ Kurkov, *The Case of the General’s Thumb* (2004) 137, translated by G. Bird. A view that Mitchell seems close to sharing about the present state of corporate America: see Mitchell, *Corporate Irresponsibility* (2001) 93-4, 258-9.

CONCEPTUALISING THE CONSTRUCTIVE TRUST

*Andrew D. Hicks, Law School, University of Hull **

Introduction

In an oft-quoted article, published in this journal in 1977, Ronald Maudsley wrote that the constructive trust “is an instrument created by the law to do justice” but that it also “needs to have some shape.”¹ Almost thirty years on and we are still trying to find that shape, or at least one that can do the job adequately. The conceptual choice is apparently two-fold: either “institutional” or “remedial”. The institutional constructive trust is characterised as a rule-based property entitlement that arises at the time of the facts giving rise to the claimant’s cause of action. The remedial constructive trust is characterised as a discretionary remedy that confers on the claimant a proprietary interest only when declared by a court of competent jurisdiction.² The choice, however, is not simply one of alternatives. Judges and jurists have not been uniform in the content they have given either abstraction. These numerous differences, not always appreciated, have potentially far reaching implications.

How, then, are we to choose between these possibilities? The constructive trust does not exist in nature. It is not capable of being simply tracked down in its natural environment and scientifically dissected to expose its essential anatomy. Rather, it is a construct developed in legal discourse to serve both

* This article draws upon doctoral research undertaken at the University of Hull. I am grateful to Nick Parry for his supervision of that project. I would also like to extend special thanks to Craig Rotherham of Nottingham University for his extensive and helpful comments on earlier drafts of this paper. My thanks also to an anonymous referee who made a number of useful suggestions. Responsibility for the final product remains with me.

¹ Maudsley, “Constructive Trusts” (1977) 28 *NILQ* 123 at p.137.

² On the distinction see *Westdeutsche Landesbank Girozentrale v Islington LBC* [1996] AC 669 at 714-715 *per* Lord Browne-Wilkinson; *Re Polly Peck International plc (in administration) (No.2)* [1998] 3 All ER 812; *Fortex Group Ltd v MacIntosh* [1998] 3 NZLR 171 at 172-173, 175-177 *per* Tipping J; *Commonwealth Reserves I v Chodar* [2001] 2 NZLR 374 at [36]-[63] (Glazebrook J); *Dickie v Torbay Pharmacy* (1986) Ltd [1995] 3 NZLR 429 at 441 (Hammond J); *Atlas Cabinet & Furniture Ltd v National Trust Co* (1990) 45 BCLR (2d) 99 at 112 *per* Lambert JA; *Muschinski v Dodds* (1985) 160 CLR 583 at 614 *per* Deane J; *Re Omegas* 16 F 3d 1443 (6th Cir. 1994); Maddaugh & McCamus, *The Law of Restitution* (2nd ed., 2004), pp.106-128; Rotherham, *Proprietary Remedies in Context: A Study in the Judicial Redistribution of Property Rights* (2002), pp.12-32; Wright, *The Remedial Constructive Trust* (1998); Martin, *Hanbury & Martin: Modern Equity* (16th ed., 2001), pp.301-304; Wright, “The Remedial Constructive Trust and Insolvency” in *Restitution and Insolvency* (Rose ed., 2000) 206; Birks, “Rights, Wrongs, and Remedies” (2000) 20 OJLS 1 at pp.9-19; Sir Peter Millett, “Restitution and Constructive Trusts” (1998) 114 LQR 399.

normative and analytical functions.³ In assessing the utility of the proposed approaches at least four considerations therefore appear relevant. First, the chosen approach should permit adequate consideration and implementation of the appropriate normative concerns. In the context of specific relief such considerations must be considered in two different contexts: the justice of the claimant's entitlement as against the defendant; and the justice of that claim as against third parties claiming through the defendant. The latter consideration is relevant even where no third parties are apparent, not least because of the risk that important third party claims may not in fact be known by the court. Second, the chosen approach should facilitate the systematic organisation and presentation of the law, thereby enabling a reasonable measure of legal certainty, coherent development and synthesis. This requires the chosen approach to be logical and analytically sound. Third, the manner in which the constructive trust is portrayed should reflect realistically what the courts are doing in fact when they hold that the defendant is a constructive trustee of specific property. Care should be taken to avoid concepts which obscure more than they reveal. Finally, any conceptual revision, if required, should be the minimum necessary to achieve the desired results. This limitation, the principle of minimal conceptual disturbance, is necessary because the impact of a particular modification in one part of a legal system may effect changes to another, whether intended or not.⁴

This article examines the various ways in which the constructive trust is conceptualised, drawing upon judicial decisions and academic comments from England, the United States and the Commonwealth. It suggests that existing approaches in these jurisdictions are, to varying degrees, found wanting when measured against the criteria identified above. The orthodox Anglo-American approach, examined in Part II, turns out to be incoherent, fits poorly with related doctrines, and obscures important and complex normative concerns. The reasons for its persistence are ideological and historical. By contrast, courts and commentators in the Commonwealth have taken a more functional approach to the constructive trust. However, none of the suggested alternatives to the Anglo-American orthodoxy are entirely satisfactory. In Part III we see that a number of Canadian judges have attempted to develop a form of *in specie* relief that avoids unfair prejudice to third parties but, nevertheless, have continued to force such relief into the constructive trust mould. This, confusingly, gives us one concept with two distinct functions. Others have suggested that the constructive trust should allow the court to grant specific relief on whatever terms, and against whichever parties, that the court sees fit. This creates a range of problems that we might expect to find whenever judges are granted strong discretion. In Part IV we then see that Australian courts and commentators have developed approaches that are either excessively illogical or premised on wholesale changes that risk unintended interference elsewhere in the legal

³ On the nature of such concepts generally, see Hart, "Definition and Theory in Jurisprudence" in *Essays in Jurisprudence and Philosophy* (1983) 21; Kelly, "Legal Concepts, Logical Functions and Statements of Fact" (1968) 3 U Tas L Rev 43; Ross, "Tu-Tu" (1957) 70 Harv L Rev 812.

⁴ See Waldron, "'Transcendental Nonsense' and System in the Law" (2000) 100 Colum L Rev 16.

system. Moreover, the approaches do not deal effectively with the very problem that motivated them; namely, the creation of unwarranted and unintended priorities.

In light of these findings I suggest in Part V that a new analytical framework is required. This new framework would recognise both the constructive trust and a purely personal order to transfer specific property. The constructive trust would be conceptualised as a judicially imposed proprietary interest operating retrospectively, from the time of the facts giving rise to the claimant's cause of action, and would be imposed only where there is reason to grant the claimant the significant advantages that follow from the recognition or creation of an equitable proprietary interest. The purely personal order to transfer specific property, by contrast, would not recognise or create an equitable interest; it would simply confer on the claimant a purely personal right to the transfer of a particular asset. The claimant would be placed in a similar position to any other unsecured creditor and would not acquire an interest in the disputed asset until the defendant complied with his duty to transfer it. This would enable the courts to award a claimant a specific asset, rather than a money substitute, without the risk of creating unintended and unwarranted priorities. By reconceptualising the constructive trust and recognising a purely personal order to transfer specific property along these lines, normatively desirable results can be achieved within a framework that is coherent, rational and illuminating.

Conceptions of the Constructive Trust as an “Institution”

(a) Anglo-american orthodoxy: the automatic vesting constructive trust

Orthodox institutional theory conceptualises the constructive trust as arising by rule of law, independently of the court order, from the moment that the claimant's cause of action arises. The function of the court is simply to establish that the relevant facts did indeed occur and to recognise the claimant's pre-existing entitlement.⁵ A claimant does not ask a court to *impose* an institutional constructive trust; rather he alleges that one “already has come into existence as a result of the way in which the defendants obtained possession of their property”.⁶ Facts occur upon which there is a simultaneous and automatic bifurcation of legal and equitable interest. The claimant's equitable proprietary interest arises once and for all, subject only to known modes of extinction, from this time.⁷ This “classical”⁸

⁵ See the authorities cited in n.2 above.

⁶ *Kuwait Oil Tanker Company SAK v Al Bader*, unreported, December 17, 1998, QBD (Moore-Bick J), reversed on other grounds [2003] UKHL 31. See also *Re Jarvis* [1958] 1 WLR 815 at 819 per Upjohn J.

⁷ See, e.g. *Grant v Edwards* [1986] Ch 638 at 652 per Mustill LJ; *Turton v Turton* [1988] Ch 542 at 552 per Nourse LJ, 554-555 per Kerr LJ; *Sen v Hedley* [1991] Ch 425 at 440 per Nourse LJ; *Attorney General for Hong Kong v Reid* [1994] 1 AC 324. Acceptance of this is typically combined with forceful rejection of the remedial constructive trust: *Re Polly Peck International (in administration) (No2)*, n.2 above; *Metall und Rohstoff v Donaldson Lufkin Inc* [1989] 3 All ER 14 at 57 per Slade LJ. But compare *England v Guardian Insurance* [2000] Lloyd's Rep IR

conceptualisation predominates in English law. More surprisingly it has also taken root in the United States, although the constructive trust in this jurisdiction is formally declared to be remedial.⁹ Scott in particular is emphatic, arguing that “[t]he beneficial interest in the property is from the beginning in the person who has been wronged. The constructive trust arises from the situation in which he is entitled to the remedy of restitution, and it arises as soon as that situation is created”.¹⁰ The role of the court, therefore, involves nothing more than the specific enforcement of the claimant’s pre-existing rights. The court order places property in the hands of the “proper and equitable owner”.¹¹

(b) The problematic consequences of the automatic vesting approach

(i) Obscuring the relative merits of competing claims

The conceptualisation of the constructive trust as autonomous, arising prior to the court’s involvement and apparently existing separate from any judicial recognition, risks suffocating substantive justice and marginalising legitimate third party interests. It encourages judges to think that the court has no control over the outcome and obscures complex normative issues that require consideration.¹²

Where the relevant factual conditions are present, it is said, a trust arises automatically “in the world beyond the court”,¹³ creating in the claimant an equitable proprietary interest. It thus follows from the nature of the interest previously created that third parties are inevitably affected, with such consequences as priority on the defendant’s insolvency appearing natural and self-evident. It is therefore easier for the court to deny responsibility for the impact that the constructive trust has on third parties as a result of the suggestive power of the notion that something real has previously come into existence, with pre-determined consequences.¹⁴

The apparent inevitability of the outcome is also bolstered by the power of property: at the time the court adjudicates on the dispute the claimant already owns the asset. The court cannot therefore order an alternative form of relief

404 (Thornton QC) (preparedness to consider a claim to a remedial constructive trust if the matter ultimately arose).

⁸ *French v Mason* [1999] FSR 597 at 603 per Pumfrey J.

⁹ The presentation of the constructive trust as automatically vesting is found in the *Restatement* and the leading texts: Scott & Seavey (Reporters), *Restatement of the Law of Restitution: Quasi-Contracts and Constructive Trusts* (1937) s.160; Scott & Fratcher, *The Law of Trusts* (4th ed., 1987 & Supplements), Vol. 5, s.462; Pomeroy, *Equity Jurisprudence* (5th ed., 1941), Vol. 2, s.375.

¹⁰ Scott & Fratcher, n.9 above, s.462.4.

¹¹ *United States v Fontana* 528 F Supp 137 at 144-145 (SD. NY. 1981). See also *Restatement*, n.9 above, s.160, comment e.

¹² See, e.g. the analysis of Browne-Wilkinson J, as he then was, in *Re Sharpe* [1980] 1 All ER 198.

¹³ Birks, “Proprietary Rights as Remedies” in *Frontiers of Liability* (Birks ed., 1994), Vol. 2, 214 at p.215.

¹⁴ See, e.g. Stone, “The Reification of Legal Concepts: *Muschinski v Dodds*” (1986) 9 UNSWLJ 63.

and refuse to recognise the claimant's right, for to do so would be tantamount to an illegitimate judicial expropriation of property. The idea that something is already the property of the claimant thus exerts greater leverage than competing normative considerations that may suggest an alternative outcome.¹⁵

An unreflective acceptance of the idea that the constructive trust has an existence autonomous from its judicial recognition, particularly when coupled with axiomatic notions about property, therefore wields a powerful ideological force. Thus, some commentators suggest that priority on insolvency by way of constructive trust follows not because it is just, fair or merited but because "equity" has previously created a property right. Absent this pre-existing property right the award of priority on the defendant's insolvency would constitute an improper preference:¹⁶ the role of the court is portrayed as one of enforcing pre-existing property rights rather than establishing priorities between the claimant and the defendant's creditors. The constructive trust claimant's priority is simply a necessary, if regrettable, by-product of the earlier rights creation.¹⁷

This reasoning tends to conceal rather than reveal the relative merits of competing claims while simultaneously explaining away the impact of the remedy on third parties.¹⁸ To say that the claimant is entitled to an asset, with priority over unsecured creditors, because a constructive trust exists, is different to saying that a constructive trust exists because a court will order specific relief and confer priority over certain other interests.¹⁹ The latter formulation makes it clear that the consequences require justification and recognises that "constructive trust is the name we give to that decision, not

¹⁵ See, e.g. Cohen, "Transcendental Nonsense and the Functional Approach" (1935) 35 Col L Rev 809 at p.820 (identifying "property" as one of the "magic 'solving words' of traditional jurisprudence"); Gray, "Property in Thin Air" [1991] CLJ 252 at p.305 (noting the "powerful and yet wholly spurious moral leverage" generated by appeals to "property"). For an illustration, see Birks, "Trusts Raised to Reverse Unjust Enrichment: The *Westdeutsche Case*" [1996] RLR 3 at p.14 (suggesting that the only good reason for a court to order assets to be transferred to the claimant is if they already belong to him).

¹⁶ See, in particular, Jennings & Shapiro, "The Minnesota Law of Constructive Trusts and Analogous Equitable Remedies" (1941) 25 Minn L Rev 667 at pp.675-676; Scott, n.9 above, s.462.4.

¹⁷ While equity may have created the property right on the basis of existing precedent, the precedent may be the product of historical accident rather than a carefully reasoned solution to the problem of priority. Constructive trusts evolved historically to resolve two-party disputes, with little overt concern or consideration for creditors' interests: Maitland, *Equity: A Course of Lectures* (1936), p.220; Waters, "The English Constructive Trust: A Look Into the Future" (1966) 19 Vanderbilt L Rev 1215 at pp.1216-1220. Scots lawyers reject the widespread adoption of the constructive trust on this ground. Why, they ask, should Scottish insolvency law be determined by the accidents of another jurisdiction? See Gretton, "Constructive Trusts" (1997) 1 Edinburgh L Rev 281 & 408 at pp.410-411.

¹⁸ Dawson, *Unjust Enrichment: A Comparative Analysis* (1951), pp.31-32; Lacy, "Constructive Trusts and Equitable Liens in Iowa" (1954) 40 Iowa L Rev 107 at pp.109-114; Sherwin, "Constructive Trusts in Bankruptcy" [1989] U Ill L Rev 297 at pp.311-313.

¹⁹ Cohen, n.1 5 above, pp.813-814.

the reason for it”.²⁰ It is therefore less likely to propagate constructive trusts that have unjust consequences on the defendant’s insolvency. Thus, where judges are asked openly to *retrospectively* impose equitable title to assets in favour of the claimant, creditors and supervening interests are at the forefront of their minds. A constructive trust is imposed, if at all, cautiously.²¹ By contrast, the first formulation appears to naturalise such consequences:²² the disputed asset already belongs to the claimant, not the defendant, therefore ordering its transfer has no impact on the defendant’s creditors.²³ It is therefore easier to lose sight of the essential policies that should limit the constructive trust’s application.

(ii) Logical deficiencies

We also find that the automatic vesting constructive trust is a somewhat illogical judicial creation. For example, in some contexts an element of discretion is formally recognised as part of the decision to award a constructive trust,²⁴ even though this is inconsistent with the automatic vesting conceptualisation. It makes no sense to characterise the interest under the constructive trust as arising automatically at the time of the facts because, if the interest is dependent on how a court exercises its discretion, it cannot arise independently of and prior to the judicial decision that creates it.²⁵ Talk of discretion might, of course, be downplayed²⁶ or passed off as

²⁰ Dobbs, *Law of Remedies - Equity - Damages - Restitution* (1993), p.399, n.40.

²¹ See, e.g. *Fortex*, n.2 above, at 179-180 *per* Henry J; *Chodar*, n.2 above, at [57] (the interests of creditors and third parties are a “major and frequent obstacle to the grant of a proprietary remedy”).

²² For a useful account of the ways in which legal concepts can make particular conclusions appear self-evident while avoiding discussion about the desirability of those conclusions, see Cohen, “The Value of Value Symbols in Law” (1952) 52 *Col L Rev* 893.

²³ See, e.g. Scott, who avoids any reference to the relevant policy concerns. Scott reasons that the constructive trust arises at the time of the unjust enrichment, and for this reason the claimant is awarded priority, before demonstrating that the claimant’s interest under a constructive trust vests automatically because the claimant can recover property *in specie* even though the wrongdoer is insolvent: Scott & Fratcher, n.9 above, s.462.4, pp.323, 325.

²⁴ E.g. where a constructive trust is imposed as the minimum equity to do justice in appropriate cases of proprietary estoppel, the determination of the minimum equity being formally portrayed as discretionary: see, e.g. *Campbell v Griffin* [2001] EWCA Civ 990 at [36] *per* Walker LJ; *Griffiths v Williams* (1977) 248 EG 947.

²⁵ A broader conceptual inconsistency is also becoming apparent as some English courts resort to the separation of liability and remedy more generally in order to determine whether a constructive trust is appropriate. They recognise that remedy is not necessarily pre-determined by the nature of the obligation breached but rather follows from a search for what is factually the better solution to the facts before the court: see, e.g. *Lord Napier & Ettrick v Hunter* [1993] AC 713 at 738 *per* Lord Templeman, 744 *per* Lord Goff (each providing reasons why a lien was a more appropriate remedy than a constructive trust, despite the existence of authority supporting the imposition of the latter). See also *Ocular Sciences Ltd v Aspect Vision Care Ltd* [1997] RPC 289 at 411-416, where Laddie J rejected explicitly the existence of a direct link between profitable breach of fiduciary obligation and constructive trust, opining that “the court should consider whether it is the appropriate remedy in the circumstances of the case”.

rhetoric.²⁷ But while a broad or strong discretion may not exist, it is not always possible or desirable to eliminate discretion entirely, even in the face of a formal commitment to rules.²⁸

A further logical deficiency is evident when one considers the link between the time at which the equitable interest apparently arises under the constructive trust and the time of the facts that determine the interest. A number of commentators suggest that logic requires the constructive trust to arise at the time of the events creating a cause of action in the claimant.²⁹ However, facts occurring after the cause of action arose may be relevant to the decision to award or deny a constructive trust.³⁰ This is true of constructive trusts arising in the context of specifically enforceable contracts for sale,³¹ even though the constructive trust is said to arise immediately upon completion of the contract.³² It is also true of estoppel cases, where supervening events may lead the court to either withhold the required relief altogether or grant relief of a different form or on different terms than it would have done if asked at an earlier stage.³³ It now appears, moreover, that similar observations might be made in the context of quantifying parties' respective beneficial entitlements under a common intention constructive trust, at least in those cases where the parties' respective shares are not clear from what was done or said at the time the common intention was formed.³⁴

²⁶ See, e.g. *Jennings v Rice* [2002] EWCA Civ 159 at [43]-[44] per Walker LJ; Gardner, "The Remedial Discretion in Proprietary Estoppel" (1999) 115 LQR 438.

²⁷ See, e.g. Edelman, "Remedial Certainty or Remedial Discretion in Estoppel After *Giumelli*" (1999) 15 JCL 179 (talk of discretion simply refers to rule-based discretion). For an alternative analysis, see Wright, "*Giumelli*, Estoppel and the New Law of Remedies" [1999] CLJ 476.

²⁸ It is increasingly recognised that the difference between rules and discretion is one of degree and somewhat unstable, varying with time and context: see Sunstein, "Problems With Rules" (1995) 83 Cal L Rev 953; Schneider, "Discretion and Rules: A Lawyer's View" in *The Uses of Discretion* (Hawkins ed., 1992) 47. Any distinction between the institutional and remedial made on the basis of rules-discretion is therefore somewhat ephemeral: see Gardner, "The Element of Discretion" in *Frontiers of Liability* (Birks ed., 1994), Vol. 2, 186 at pp.193, 197-198.

²⁹ See, e.g. Kull, "Restitution in Bankruptcy: Reclamation and Constructive Trust" (1998) 72 American Bankruptcy LJ 265 at p.287; Black, "*Baumgartner v Baumgartner*, the Constructive Trust and the Expanding Role of Unconscionability" (1988) 11 UNSWLJ 117 at pp.128-129.

³⁰ Although this is often denied: see, e.g. *Turton v Turton* [1988] 1 Ch 542 at 552 per Nourse LJ.

³¹ See, e.g. *Price v Strange* [1978] 1 Ch 337.

³² *Lysaght v Edwards* (1876) 2 Ch D 499 at 506-507 per Lord Jessel MR, although as to the precise nature of the interest arising see *Jerome v Kelly (Inspector of Taxes)* [2004] UKHL 25, [2004] 1 WLR 1409 at [30]-[32].

³³ For particularly notable examples, see *Crabb v Arun DC* [1976] Ch 179 at 189-190 per Denning MR, 199 per Scarman LJ; *Williams v Staite* [1979] Ch 291 at 298-299 per Goff LJ, 300-301 per Cumming-Bruce LJ; *Pascoe v Turner* [1979] 1 WLR 431; *Burrows v Sharp* (1989) 23 HLR 82; *Voyce v Voyce* (1991) 62 P & CR 290 at 296 per Nicholls LJ; *Sledmore v Dalby* (1996) P & CR 196.

³⁴ As Chadwick LJ noted recently, "there is no difference, in cases of this nature, between constructive trust and proprietary estoppel": *Oxley v Hiscock* [2004] EWCA Civ 546, [2004] 3 WLR 71 at [66]. See also *Grant v Edwards* [1986] Ch 638 at 656 per Browne-Wilkinson VC; *Yaxley v Gotts* [2000] Ch 162 at 176, 180

In these instances the court will grant a share “which, in light of all the material circumstances (including the acts and conduct of the parties after the acquisition) is shown to be fair”.³⁵ Thus, each party will be:

“entitled to that share which the court considers fair having regard to the whole course of dealing between them in relation to the property. And in that context, ‘the whole course of dealing between them in relation to the property’ includes the arrangements which they make from time to time in order to meet the outgoings (for example, mortgage contributions, council tax and utilities, repairs, insurance and housekeeping) which have to be met if they are to live in the property as their home.”³⁶

It follows that the claimant’s beneficial entitlement will depend, in part at least, on facts occurring after the date on which the common intention was formed and detrimentally relied upon. It is, however, the date on which the common intention was detrimentally relied upon that the automatic vesting approach fixes the claimant’s beneficial interest.³⁷

We find, therefore, that in many cases facts which determine the availability or extent of the claimant’s beneficial entitlement under a constructive trust can occur at a date later than its apparent date of birth. In these instances at least, logic suggests that the constructive trust must exist only from the date at which the factors determining the existence or extent of the beneficial interest crystallise; namely, at the date of court hearing. Any equitable proprietary interest “existing” from the time the claimant’s cause of action accrued must be recognised as created retrospectively.

We also find problems where the constructive trust is contingent on the exercise of a power of election by the claimant. Thus, where the claimant

per Walker LJ; *Chan Pui Chun v Leung Kam Ho* [2002] EWCA Civ 1075 at [91]-[92] *per* Parker LJ. The apparent coalescence is not, however, universally accepted: see Fox, “Trusts of the Family Home: The Impact of *Oxley v Hiscock*” (2005) 56 NILQ 83 at 99-102.

³⁵ *Oxley v Hiscock* [2004] EWCA Civ 546 at [66] *per* Chadwick LJ. Earlier cases had suggested that post-acquisition events were relevant because they served simply to demonstrate what shares must have been intended by the parties at the outset, but they did not themselves create or determine the entitlement: see *Midland Bank v Cooke* [1995] 4 All ER 562 at 574 *per* Waite LJ; *Stokes v Anderson* [1991] 1 FLR 391 at 400 *per* Nourse LJ.

³⁶ *ibid.*, at [69] *per* Chadwick LJ. See also *Grant v Edwards* [1986] Ch 638 at 657-658 *per* Browne-Wilkinson VC; *Drake v Whipp* [1996] 1 FLR 826 at 831 *per* Gibson LJ; *Cox v Jones* [2004] EWHC 1486, [2004] 2 FLR 1010; *Hurst v Supperstone* [2005] EWHC 1309.

³⁷ In *Oxley v Hiscock*, Chadwick LJ did not discuss the impact of his conclusions on the nature of the constructive trust, but recognised at [70] that “the courts have not found it easy to reconcile [the approach] with a traditional property-based approach.” Elsewhere, it has been noted that an estoppel-based approach, involving the imposition of a constructive trust with retrospective effect, “is unlikely to find favour with English courts whose approach is to regard rights of property as fixed and ascertainable in advance and immutable”: Hayton, *Commentary & Cases on the Law of Trusts and Equitable Remedies* (11th ed., 2001), p.352.

seeks a constructive trust pursuant to the successful tracing of an asset into a substitute, situations arise in which the original asset remains identifiable in the hands of a third party, who took with notice, while the substitute is identifiable in the hands of the defendant. The claimant must therefore elect against whom he wishes to proceed. He cannot simultaneously own both the original asset and the substitute.³⁸ Prior to the election, the claimant cannot therefore have a full equitable interest in the substitute asset. How, then, can a constructive trust “arise” prior to the date of election? The same question can be asked of cases where property passes pursuant to a voidable contract. Failure to rescind precludes a constructive trust claim, and the consequent equitable proprietary relief,³⁹ on the ground that despite the impediment affecting the transfer the transferor nonetheless intended full legal and beneficial ownership to pass.⁴⁰ The constructive trust cannot therefore “arise” at the time of the transfer⁴¹ because its existence is contingent on the subsequent avoidance by the transferee. The transferee is only “potentially a constructive trustee”.⁴² If the equitable interest that arises pursuant to rescission is to have effect from the time of the initial transfer, this must be recognised as being retrospective.⁴³

(iii) *Gross artifice*

Elsewhere, we find that explanations demanded by the automatic vesting approach, while logical, do not appear to reflect the reality. For example, where the defendant acquires a benefit from a third party in breach of fiduciary obligation, the principal may elect to adopt or reject the fiduciary’s purchase.⁴⁴ This appears inconsistent with the idea of a beneficial interest vesting automatically in the claimant upon the defendant’s receipt because the claimant’s interest is contingent upon a positive election. We are, however, told by the authorities that, upon the defendant’s acquisition, the claimant immediately acquires a full equitable interest. Where the election is

³⁸ Birks, “Mixing and Tracing” (1992) 45 CLP 69 at pp.89-98, developing more fully the views expressed in *An Introduction to the Law of Restitution* (1989, revised edition), pp.91-93, 393-394. Smith suggests simultaneous ownership is possible: *The Law of Tracing* (1997), pp.380-383. But see the compelling criticism of this view by Rotherham, n.2 above, pp.95-96.

³⁹ See *Twinsectra v Yardley* [1999] Lloyd’s Rep Bank 438 at 461 *per* Potter LJ; *Halifax Building Society v Thomas* [1996] Ch 219 at 226 *per* Gibson LJ; *El Ajou v Dollar Land Holdings* [1993] 3 All ER 717 at 734 *per* Millet J; *Lonrho v Fayed (No 2)* [1992] 1 WLR 1 at 11-12 *per* Millett J.

⁴⁰ *Collings v Lee* [2001] 2 All ER 332 at 337 *per* Nourse LJ; *Halley v Law Society* [2003] EWCA Civ 97 at [53] *per* Carnwath LJ.

⁴¹ As claimed in *Latec Investments v Hotel Terrigal Pty Ltd* (1965) 113 CLR 265 at 290-291 *per* Menzies J (in the event of successful rescission “the conveyee holds, and has always held, as trustee”); *Stump v Gaby* (1852) 2 De G M & G 623. See also Cope, *Constructive Trusts* (1992), pp.278-280.

⁴² *Estates Realities v Wignall* [1992] 2 NZLR 615 at 632 (Tipping J).

⁴³ See *Lonrho v Fayed (No 2)*, n.39 above, at 12. There, Millett J noted that it may well be “the beneficial interest in the property will be treated as having remained vested in him throughout” (emphasis added).

⁴⁴ See *Dean v MacDowell* (1878) LR 8 Ch D 345 at 351 *per* James LJ. As a leading text puts the matter, in such cases the constructive trust is “a sword for the beneficiaries which may be sheathed if desired”: Mowbray, Tucker, Le Poidevin & Simpson, *Lewin on Trusts* (17th ed., 2000), p.447.

positive the claimant simply takes that which, in equity, is already his; where the election is negative that act simply involves the making of a gift of the beneficial interest to the defaulting fiduciary.⁴⁵ This is surely a perversion of words, if not intellectually dishonest. The “gift” is really nothing of the sort; it is merely a condonation of a breach of fiduciary duty.

(iv) Restricting the availability of relief in specie

The automatic vesting approach may also impact negatively on the general availability of specific relief. An important feature of the constructive trust is that it allows the claimant to take the very property that is the subject matter of the dispute. He is not left with an equivalent monetary substitute. Another important feature is that the constructive trust confers priority over the defendant’s creditors and later equitable interests. The result of this is that, on occasion, the courts may refuse to extend specific relief by way of constructive trust to encompass the case at hand, where third parties are absent, for fear of binding a court hearing a future case, where third party interests are present.⁴⁶ The problem arises because the constructive trust is assumed to create equitable property rights which arise necessarily at the time of the facts that fulfil the necessary conditions for its existence. Third parties claiming through the defendant are therefore bound of necessity whenever the relevant triggering facts are established. Consequently, the availability of specific relief outside the insolvency context may be limited, despite reasons other than priority for its award,⁴⁷ for fear of forcing future courts to recognise the existence of inappropriate rights. While the problem could be avoided if the courts were prepared to take a more discretionary approach to the constructive trust, retaining the power to deny its availability where it would create an injustice, this would be inconsistent with the orthodox rejection of discretionary constructive trusts in English law.⁴⁸ We thus find a lacuna in equity’s remedial armoury.

(c) Why has the automatic vesting approach persisted?

The automatic vesting conceptualisation has been rarely questioned in English law and, despite recent challenges,⁴⁹ continues to represent the orthodoxy in the United States.⁵⁰ Given the logical and practical deficiencies

⁴⁵ For particularly strong reliance on this theory, see *Cook v Evatt (No 2)* [1991] 1 NZLR 676 at 696-697. The theory also appears in *Cook v Deeks* [1916] 1 AC 554 at 564, where the Privy Council established that the self-interested directors had no power to condone their breach and “make a present to themselves” of company property.

⁴⁶ See, e.g. Klippert, *Unjust Enrichment* (1983), p.196 (priority advantage conferred by the constructive trust means that it cannot be made broadly available for fear of the incidental creation of unwarranted priorities).

⁴⁷ On which see Goff & Jones, *The Law of Restitution* (5th ed., 1998), pp.76-77.

⁴⁸ See also the criticisms of the “limited flexibility” and “monolithic” remedial approaches below.

⁴⁹ See, e.g. *Re CRS Steam Inc.* 225 BR 833 (Bankr. Mass. 1998); *Re Dow Corning Corp* 192 BR 428 (Bankr. Mich. 1996); *US v BCCI Holdings (Luxembourg), SA* 46 F 3d 1185 (DC Cir. 1995); *Re Omegas Group Inc* 16 F 3d 1443 (6th Cir. 1994).

⁵⁰ See, e.g. *Re McCafferty* 96 F 3d 192 at 198 (6th Cir. 1996); *Re Dameron* 115 F 3d 718 (4th Cir. 1998); *Re Leitner* 236 BR 420 (Bankr. D. Kan. 1999).

of this conceptualisation its persistence appears puzzling. Three reasons might explain its longevity.

(i) Obscuring the judicial variation of property rights

One explanation for the persistence of the automatic vesting approach is its use in obscuring departures from classical liberal understandings of property. The proposition that the judiciary has no power to engage in the non-consensual redistribution of property rights permeates much of English law.⁵¹ However, some constructive trusts are incompatible with this orthodoxy. For example, the constructive trust of the bribe in *Attorney General for Hong Kong v Reid*⁵² was redistributive: it did not arise as a result of the consent of the parties and it did not function to protect a pre-existing property right of the claimant. However, rather than acknowledge this the Privy Council reasoned that the constructive trust arose automatically upon receipt of the bribe, prior to the involvement of the court, by virtue of the maxim “equity considers as done that which ought to be done”. The court did no more than simply recognise the claimant’s pre-existing right. It thereby formally maintained the appearance that orthodox notions of property were not being violated while creating in fact a property right in the claimant *de novo*.⁵³ The automatic vesting conceptualisation thus suppresses a reality that is subversive to the basic commitments of English law. Its function, as Rotherham points out, is “not so much practical as ideological”.⁵⁴ It is, therefore, no coincidence to find movement away from the automatic vesting conceptualisation in those jurisdictions which have developed more instrumentalist understandings of property. In Canada, for example, it is now accepted that the courts can both recognise and create property rights by the imposition of a constructive trust.⁵⁵

(ii) The ideal of pre-legal rights

The portrayal of the constructive trust as an automatically vesting entitlement also maintains congruence with the classical idea of pre-legal rights that derive from some privileged source of legitimacy. These ideals underpin the work of Professor Birks: the role of the judge is to find and apply the law, changing nothing except that which comes from a better understanding of what is there already.⁵⁶ With such broader methodological commitments the automatic vesting conceptualisation of the constructive trust is perhaps

⁵¹ Rotherham, n.2 above, pp.34-40

⁵² [1994] 1 AC 324.

⁵³ Rotherham, “Proprietary Relief for Enrichment by Wrongs: Some Realism About Property Talk” (1996) 19 UNSWLJ 378 at pp.396-397. See also *Cook v Deeks*, n.45 above (opportunity acquired by directors from a third party in breach of fiduciary duty characterised as usurpation of the company’s property).

⁵⁴ Rotherham, n.2 above, p.29.

⁵⁵ See, e.g., *Lac Minerals v International Corona Resources Ltd* (1989) 61 DLR (4th) 14 at 50-51 *per* La Forest J; *Semiahmoo Indian Band v Canada* (1997) 148 DLR (4th) 523 at 559-560 *per* Isaac CJ.

⁵⁶ See, e.g. Birks, *An Introduction to the law of Restitution* (1989, revised edition), pp.23, 27 and “The End of the Remedial Constructive Trust?” (1998) 12 TLI 202 at pp.214-215. The similarities to the heavily criticised classical legal thought of late nineteenth century America are numerous: see, e.g. Grey, “Langdell’s Orthodoxy” (1983) 45 U Pitt L Rev 1.

inevitable. A constructive trust that arises automatically, prior to any judicial involvement, and which merely recognises the claimant's pre-existing entitlement, maintains congruence with the idea that the result follows from the application of a body of impartial universal principles established since time immemorial. It thus creates the impression that the court undertakes a passive declaratory role rather than an active and creative one.⁵⁷

(iii) *The distorting influence of Scott and the Restatement*

It would, however, appear that these reasons do not explain adequately the longevity of the automatic vesting approach in the United States. There, the realist onslaught of the late nineteenth and early twentieth centuries led to the emergence of an instrumental concept of law and property which make the persistence of the automatic vesting conceptualisation superficially puzzling.⁵⁸ But such persistence does not defy all explanation. While classical visions of property and the role of the judiciary were historical stimuli for the conceptualisation, its persistence has for some time owed more to the powerful stabilising effect of the *Restatement of Restitution* and Scott's *Law of Trusts*, both of which have played a pivotal role in American understandings of the constructive trust.⁵⁹

Scott was a joint reporter of the *Restatement*, published in 1937, and sole author of the first edition of his influential treatise.⁶⁰ Both works conceptualise the constructive trust as an automatically vesting entitlement, arising prior to the involvement of the court. This is partially explicable by Scott's classical orthodox background. Scott wrote towards the end of an era characterised by a firm belief in the inviolability of property, universal principles of common law and the felt need to reveal the law's hidden

⁵⁷ Note, however, that the declaratory theory of law has been killed off judicially: *Kleinwort Benson v Lincoln City Council* [1999] 2 AC 349.

⁵⁸ On the emergence of an instrumental concept of law and its stimuli, see generally Horwitz, *The Transformation of American Law 1870-1960: The Crisis of Legal Orthodoxy* (1992). In outline, the instrumentalist approach involved: (i) a rejection of the belief that cases could be decided without controversy by reference to abstract legal principles and concepts; (ii) the recognition that judges made rather than discovered law; and (iii) the recognition that judge-made law was an instrument of social policy, to be made and applied purposively in accordance with such policies. It was equally accepted that property rights were necessarily shaped and limited by judges balancing competing policies and interests, with each stick in the owner's bundle of rights necessarily imposing burdens on non-owners. See, e.g. Singer, "Legal Realism Now" (1988) 76 Calif L Rev 467; Horwitz, Fisher & Reed, *American Legal Realism* (1993); Cohen, "Property and Sovereignty" (1927) 13 Cornell LQ 8; Hohfeld, "Some Fundamental Legal Conceptions as Applied in Judicial Reasoning" (1913) 23 Yale LJ 16; Hohfeld, "Fundamental Legal Conceptions as Applied in Judicial Reasoning" (1917) 26 Yale LJ 710.

⁵⁹ Both have been cited heavily for the automatic vesting approach: see the cases cited in Scott & Fratcher, n.9 above, s.462.4. Scott in particular is often cited as authority where the effect of timing is to confer priority over the defendant's creditors or intervening lien holders: see, e.g. *Re General Coffee Corp* 828 F 2d 699 (11th Cir. 1987) (insolvency); *United States v Fontana*, n.11 above (federal tax lien).

⁶⁰ Scott, *The Law of Trusts* (1st ed., 1939).

structure through conceptual ordering.⁶¹ Moreover, the formative years of his legal education were spent at Harvard, the heartland of Langdellian casebook method which taught that the law could be discovered scientifically and arranged schematically as a set of logical propositions.⁶² Equally, the early Restatements represented an instinctive reaction by the legal establishment to the realist movement and its criticism of Langdellian legal science. They attempted to refute claims about the legislative role of the judge while simultaneously reasserting the idea of impartial, self-executing laws.⁶³ A conceptualisation that bolstered these ideals was therefore understandable. The problem is that Scott's treatise and the *Restatement* have been perceived as authorities of timeless value, crystallising the cases into properly derived rules that are simply there to be applied. Thus, it has been noted that restatements of law were often treated in a similar way to legislative codes. So long as authority did not lie squarely against the rules they promulgated, they could be followed by judges "exactly as they would follow a statute".⁶⁴ Much the same might be said of Scott's treatise: it became almost too authoritative to challenge. Consequently, both authorities continued to exert an influence that retarded change in light of new understandings about the nature of property, the judicial function, and the efficacy and fairness of equitable property rights on insolvency.⁶⁵ In the United States, the persistence of the automatic vesting conceptualisation is therefore illustrative of a broader problem created by the treatise tradition and restatements of law.

(d) The constructive trust as automatically retrospective in effect

While rejecting the notion of the remedial constructive trust Tony Oakley, author of the leading English text on the subject, has formulated an alternative conceptualisation that differs from the orthodoxy in at least one fundamental respect: a court order is necessary for the constructive trust to take effect.⁶⁶ He thus talks of the "imposition" rather than "recognition" of a constructive trust. This conceptualisation sits a little uneasily with the idea of institutional trusts. However, Oakley openly rejects discretionary remedial approaches⁶⁷ and views the consequences of the constructive trust as monolithic.⁶⁸ Moreover, he characterises the interest of a potential constructive trust claimant who has yet to secure a declaration of his interest

⁶¹ Rotherham, n.2 above, pp.49-51, 57-59. See also Grey, n.56 above.

⁶² See Waters, "The Role of the Trust Treatise in the 1990s" (1994) 59 *Mo L Rev* 121 at p.124.

⁶³ Gilmore, *The Death of Contract* (1974), pp.58-59.

⁶⁴ Gilmore, "Legal Realism: Its Cause and Cure" (1961) 70 *Yale LJ* 1037 at p.1044.

⁶⁵ Although see the more enlightened contemporary opinion of the constructive trust as a priority creating mechanism: *Re Stotler & Co* 144 BR 385 at 388 (ND. Ill. 1992); *Re Morken* 182 BR 1007 (Bankr. D. Minn. 1995); *Foos* 183 BR 149 (Bankr. ND. Ill. 1995); *Re North American Coin & Currency Ltd* 767 F 2d 1573 (9th Cir. 1985). See also the authorities cited in n.49 above.

⁶⁶ Oakley, *Constructive Trusts* (3rd ed., 1997), pp.5-6.

⁶⁷ Oakley also appears on occasion to suggest that the equitable interest arises automatically and a court order is simply a formality: see n.69 below.

⁶⁸ Although this may be subject to limited exceptions: see n.73 below.

as “a full equitable interest”.⁶⁹ While consistent with the case law, this appears to be a restatement of the automatic vesting approach and inconsistent with the requirement of a court order before the constructive trust can take effect: a constructive trust that is judicially imposed surely cannot exist prior to its imposition.

The view that Oakley purports to adopt derives from Bogert’s treatise on American trust law.⁷⁰ However, in this work Bogert flatly denies that the claimant holds the beneficial interest from the outset. He suggests that when constructive trust claimants seek the imposition of a constructive trust, “they may be treated as if they had been from the beginning the owners of such estate or interest.”⁷¹ Thus, the constructive trust is fully retrospective; for all practical purposes it is *deemed* to have arisen at the time of the facts giving rise to its imposition.⁷² It follows that, while a court order is necessary for a constructive trust to take effect, this in no way affects the priorities between the constructive trust claimant and third parties.

(e) The utility of automatic retrospectivity in English law

While the requirement of curial declaration for a constructive trust to operate avoids reifying the constructive trust in the sense of treating it as something that exists independently of the decision to impose it, the approach continues to treat the claimant’s priority over general creditors as an inevitable incident of a constructive trust claim. Once the requisite operative facts are established the requirement of a court order is nothing more than a formality.⁷³ Oakley thus states that it is *because* the constructive trust takes effect at the moment of the conduct giving rise to its imposition, and hence creates an equitable proprietary interest at this time, that it is binding on the

⁶⁹ Oakley, “Proprietary Claims and Their Priority on Insolvency” [1995] CLJ 377 at p.424; Oakley, *Parker & Mellows: The Modern Law of Trusts* (8th ed., 2003), p.833.

⁷⁰ Bogert & Bogert, *The Law of Trusts and Trustees* (2nd ed., 1978), s.472.

⁷¹ *ibid.* In the same section Bogert asserts that the court order decreeing the constructive trust establishes the defendant as constructive trustee “as of the date of his wrongful acquisition”.

⁷² This conceptualisation has received some judicial support in the United States: *Re Dow Corning*, n.49 above, at 436; *Bly v Gensmer* 386 NW 2d 767 at 769 (Minn. App. 1986); *Healy v CIR* 345 US 273 at 282-283 (1952); *International Refugee Organization v Maryland Dry Dock Co* 179 F 2d 284 at 287 (4th Cir. 1950); *Stoehr v Miller* 296 F 414 at 426-427 (2nd Cir. 1924).

⁷³ Oakley does allude to the fact that the court has power to vary the general rule with respect to timing: “The Precise Effects of the Imposition of a Constructive Trust” in *Equity and Contemporary Legal Developments* (Goldstein ed., 1992) 427 at pp.437-438. This is difficult to reconcile with Oakley’s rejection of the remedial constructive trust and his support for the institutional approach. However, Oakley views this power as of limited scope. For example, it is not deemed appropriate in the context of constructive trusts imposed on profiting fiduciaries. Indeed, it appears that Oakley envisages the power to be limited to family property disputes and novel cases. He thus appears to view it as consistent with institutional theory because the limited exceptions “merely re-emphasise the universality of the general rule”: n.66 above, p.6.

trustee in bankruptcy.⁷⁴ Priority over creditors is thus portrayed as natural, inevitable and apparently uncontroversial. That the constructive trust takes effect at the time of the conduct giving rise to its existence is, however, a conclusion, not a justification. Rather than enter into discussion of the merits of priority, Oakley thus adopts circuitous reasoning to explain away the impact of the constructive trust on the defendant's insolvency.

The dominant motive behind the conceptualisation appears to be a concern over taxation: if a constructive trust exists irrespective of whether the beneficiary ever seeks a court order to compel the defendant to convey the property, the claimant may be subject to tax liability for property which he has never claimed and never enjoyed.⁷⁵ However, to this extent the emphasis on timing of creation appears an unnecessary complexity and only partially successful. While the approach may prevent the liability to tax of an unsuspecting claimant who has not sought to enforce a potential constructive trust claim, it appears that the constructive beneficiary who does enforce will become liable to tax for the period in which he has not enjoyed the property: once enforced the constructive trust is effective from the date of the relevant conduct. The claimant is therefore deemed to have owned the property in the interim and, as the beneficial owner, is liable to tax. The only way around this would be to make a sensible policy decision not to tax the claimant for such a period. This does not depend on the precise conceptualisation adopted.⁷⁶ In the United States, for example, the identification of money as taxable income has been held to turn on whether the recipient has "such control over it that, as a practical matter, he derives readily realizable economic value from it".⁷⁷ The constructive trust theory adopted by the court therefore "in no way determines the allowability of the claim".⁷⁸ A similar approach was adopted in Canada at a time when the automatic vesting approach dominated.⁷⁹ More recently, it was established that where the developing remedial constructive trust would frustrate statutory provisions or create practical problems it would be treated as not within the legislative intent of the relevant taxation provision.⁸⁰

Oakley's willingness to adapt the constructive trust in light of the apparent injustices that would otherwise be created in the context of taxation is telling. The most pressing, important and directly relevant concern for the

⁷⁴ Oakley, n.66 above, p.5. Presumably, for this reason, a constructive beneficiary who seeks to enforce his claim to the property will also become liable to tax as of that date, even though he has not enjoyed the benefit of the property in the interim.

⁷⁵ See Oakley, n.73 above, p.437; Oakley, n.66 above, p.6.

⁷⁶ See, e.g. Glover, "Taxing the Constructive Trustee: Should a Revenue Statute Address Itself to Fictions?" in *Trends in Contemporary Trust Law* (Oakley ed., 1997) 315.

⁷⁷ *James v US* 81 S Ct 1052 at 1055 (1960); *Rutkin v US* 343 US 130 at 137 (1952).

⁷⁸ *In the Matter of Diversified Brokers* 355 F Supp 79 at 89 (ED, Miss. 1973). See also *Wood v US* 863 F 2d 417 at 419 (5th Cir. 1989). Constructive trustees have therefore been liable to tax for the period in which they had dominion and control over the property, regardless of the constructive trust theory adopted: see *Healy v CIR* 345 US 278 at 282-283 (1952) (automatic retrospectivity); *First National Bank of Miami v US* 235 F Sup 331 (SD. Fla. 1964) (automatic vesting theory).

⁷⁹ *R v Poynton* (1972) DTC 6329.

⁸⁰ *Karavos v Canada* (1995) 57 ACWS (3d) 876; *Nelson v NMR* (1991) DTC 37 at 43.

constructive trust is not taxation but the extent to which its imposition justifiably affects innocent third parties. However, it is this very issue that Oakley strains to avoid, reflecting the traditional reluctance of English law to engage openly in this question.⁸¹

Conceptions of the Constructive Trust as a “Remedy”

(a) Judicial conceptions of the constructive trust as a remedy in Canada

Consideration of the remedial constructive trust in Canada takes place in the context of two general developments. First, the courts have increasingly recognised that the constructive trust is “both blunt and powerful with the result that important third party interests can be too easily compromised”.⁸² In particular, there has arisen considerable disquiet over the impact of the constructive trust on creditors⁸³ and purchasers of the disputed property or an interest therein.⁸⁴ Second, following the development of a general cause of action in unjust enrichment, the courts have separated the consideration of liability from that of remedy.⁸⁵ Once liability is established, the court is required to make a context specific evaluation of the most appropriate response to the unjust enrichment in question, taking into account the precise facts of the case as they present themselves at the date of the court hearing.⁸⁶ Following these developments there has been a lack of unanimity about the most appropriate way to conceptualise the constructive trust.

In *Rawluk v Rawluk* Cory J, speaking for the majority, expressed complete agreement with Scott’s automatic vesting approach and opined that “a property interest arising under a constructive trust can be recognised as having come into existence not when the trust is judicially declared but from the time when the unjust enrichment first arose”.⁸⁷ However, as we will see, Cory J appears to contradict this statement later in his judgment. Moreover, a continuing commitment to the automatic vesting approach is difficult to

⁸¹ See, e.g. Sir Peter Millett, “Remedies: The Error in *Lister & Co v Stubbs*” in *Frontiers of Liability* (Birks ed., 1994), Vol.1, 51 at p.52 (“Either the plaintiff is entitled to a proprietary remedy or he is not. If he is then the insolvency of the defendant is not a sufficient reason for withholding it from him.”)

⁸² *Re 512760 Ontario Inc* (1992) 91 DLR (4th) 719 at 731 (Adams J). See also *LeClair v LeClair* (1998) 159 DLR (4th) 638 at 651 per Ryan JA.

⁸³ See, e.g. *Soulos v Korkontzilas* (1997) 146 DLR (4th) 214 at 227, 230 per McLachlin J; *Bedard v Schell* (1987) Sask R 71 at 74-75 (Gerein J); *CDIC v Principal Savings & Trust Co* (1998) 224 AR 331 at 337-341 (Belzil J).

⁸⁴ See, e.g. *Rawluk v Rawluk* (1990) 65 DLR (4th) 161 at 188, 191 per McLachlin J, exhibiting particular concern over the creation of invisible and unregistered or unregistrable interests in land.

⁸⁵ *Sorochan v Sorochan* (1986) 29 DLR (4th) 1 at 7 per Dickson CJ. See Waters, “Liability and Remedy: An Adjustable Relationship” (2001) 64 Sask L Rev 429.

⁸⁶ A similar remedial approach is adopted for wrongs: see *Cadbury Schweppes v FBI Foods* [1999] SCR 142; *Lac Minerals*, n.55 above, at 50-52 per La Forest J.

⁸⁷ n.84 above, at 176, citing Oosterhoff & Gillese, *Text, Commentary and Cases on Trusts* (3rd ed., 1987), p.579 and Scott & Fratcher, n.9 above, s.462.4. This conceptualisation does not appear to have been entirely ruled out in later cases: see *Ellingsen v Hallmark Ford Sales Ltd* (2000) 190 DLR (4th) 47 at 70 per Lambert JA.

reconcile with the increased concern for subsequent third party interests. This concern appears to require a power to deny a constructive trust where, all other things being equal except the appearance of third party interests, a constructive trust would have been awarded.⁸⁸ According to the automatic vesting approach, however, the subsequent appearance of third party interests is an irrelevant consideration; third parties are bound necessarily by the earlier equitable interest created by the constructive trust.⁸⁹ The approach is also difficult to reconcile with the separation of liability and remedy: how can the constructive trust arise concurrently with liability if factors that determine its appropriateness, including those arising in the interim, are the subject of discretionary consideration as at the date of trial?

Further interpretations of *Rawluk*, consistent with remedial thinking, are possible.⁹⁰ Following his apparent agreement with Scott, Cory J quoted with approval the dictum of Lord Denning MR that a constructive trust “may arise at the outset, when the property is acquired, or later on, as the circumstances may require”.⁹¹ In light of this, he suggested that “even if it is declared by a court after the parties have already separated, a constructive trust can be *deemed* to have arisen when the duty to make restitution arose”.⁹² This appears to suggest that the constructive trust arises only when declared but the date from which it operates is a different matter. Two possible alternatives have emerged.

(i) Limited flexibility: prospective or retrospective operation

We might interpret the words of Cory J in *Rawluk* as indicating that, once a constructive trust is identified as the most appropriate remedy, it may be given either retrospective or prospective effect.⁹³ Where a constructive trust is declared to operate retrospectively it will take effect from the moment that the claimant’s cause of action accrued, and the consequences that follow from the creation of an equitable proprietary interest will run from this time. Where the constructive trust is declared to operate prospectively the

⁸⁸ See, e.g. *Peter v Beblow* (1993) 101 DLR (4th) 621 at 640, where Cory J noted that where the constructive trust would unfairly affect *bona fide* third party interests it would not be awarded.

⁸⁹ It might be argued that the constructive trust can vest automatically but the court may not enforce the trust if to do so would create an inequitable result, such as the creation of unwarranted priorities: *Shearer v Barnes* 118 Minn 179 at 188 (1912). For criticism of such an approach, see the discussion of the merits of *Muschinski v Dodds* (1985) 160 CLR 583, below.

⁹⁰ For a further inconsistent interpretation, see *Stirling v Buck* (1993) ACWS 1293, where Murphy J interpreted Cory J’s comments as meaning that the constructive trust “arises at the time the unjust enrichment arises but the court, in declaring such a trust, can declare it effective either from the date it arose or from the date of the court order”. This continues to vest the constructive trust with institutional characteristics and is akin to the narrow interpretation of *Muschinski v Dodds*, considered below.

⁹¹ *Hussey v Palmer* [1972] 3 All ER 744 at 747. On the mistaken interpretation of this *dicta*, see below.

⁹² n.84 above, at 176 (emphasis added).

⁹³ The words of Cory J were that the constructive trust “*can* be deemed to have arisen when the duty to make restitution arose”. They were not that it *will* be deemed to have arisen when that duty arose.

claimant's equitable proprietary interest will take effect only from the date of judgment. This interpretation appears consistent with the influential dissent in *Rawluk* by McLachlin J, who commented:

“When the court declares a constructive trust, at that point the beneficiary obtains an interest in the property subject to the trust. That property interest, it appears, may be taken as extending back to the date when the trust was ‘earned’ or perfected. In *Hussey v Palmer*... Lord Denning postulated that the interest may arise at the time of declaration or from the outset, as the case may require.”⁹⁴

In fact, all that Lord Denning postulated in *Hussey*⁹⁵ was that the relevant conduct which makes it unconscionable for a defendant to retain property may occur post-acquisition as well as pre-acquisition.⁹⁶ He made no suggestion that the Court could vary the time at which the interest arises. However, for present purposes the correct interpretation of *Hussey* is not as important as the meaning to which McLachlin J attributed it. She was clearly of the opinion that *Hussey* supported the inference that an unjust enrichment constructive trust⁹⁷ would arise only if and when declared, and then operate either retrospectively or prospectively.

(ii) Monolithic operation: deemed existence from time of unjust enrichment

Some courts view *Rawluk* as establishing the narrower proposition that, while the court must first determine whether the imposition of a constructive trust is appropriate, once imposed it operates *as if* it had arisen at the time the claimant's cause of action arose.⁹⁸ Thus, in *Rawluk* Cory J said:

“If the court is asked to grant such a remedy and determines that a declaration of constructive trust is warranted, then the proprietary interest awarded pursuant to that remedy will be deemed to have arisen at the time when the unjust enrichment first occurs.”⁹⁹

This conceptualisation was adopted by Ryan JA in *LeClair v LeClair*.¹⁰⁰ By his will J devised his estate equally between his wife, T, and his son, R. An apartment held solely by J but maintained and improved by T comprised 90% of the value of the estate. After J's death T claimed that her financial and non-financial contributions entitled her to an interest in half of the

⁹⁴ n.84 above, at 185.

⁹⁵ [1972] 3 All ER 744.

⁹⁶ A view consistent with the judgment of Lord Diplock in *Gissing v Gissing* [1971] AC 886 at 905-906. Elsewhere, Lord Denning was of the opinion that the constructive trust arose upon the facts that so affected the defendant's conscience that he could not take free of the claimant's rights. See, for example, *Binions v Evans* [1972] Ch 359 at 368.

⁹⁷ Her comments were explicitly limited to the constructive trust developed as a response to unjust enrichment: n.84 above, at 185.

⁹⁸ See, e.g. *King v Harris* (1995) NBR (2d) 161; *Barnebe v Touhey* (1994) OR (3d) 370.

⁹⁹ n.84 above, at 177.

¹⁰⁰ (1998) 159 DLR (4th) 638.

apartment by way of constructive trust and that this share did not therefore form part of J's estate at his death. It was, however, accepted that J would have drawn his will differently had he been aware of T's claim. If, at his death, J was therefore found to have held a half share in the apartment on constructive trust for T, T would have received an unintended windfall to the detriment of the innocent R. Ryan JA held that, while an unjust enrichment claim was established, the apartment was not automatically subject to a constructive trust prior to J's death. The court first had to determine that a constructive trust was appropriate and in this case it was not, both for the reasons stated and because the unjust enrichment had been appropriately remedied by the terms of the will which conferred on T a half share in the estate. Ryan JA stated:

“Cory J [in *Rawluk*] addressed the question whether, once a court declares a constructive trust, the interest arises at the time of judicial declaration, or as of the time of the unjust enrichment. His answer was that it arises at the time of the unjust enrichment. However, contrary to the appellant's argument in this case, this passage does not negate the need for a court to first determine whether an unjust enrichment occurred and whether the appropriate remedy would be a constructive trust.”¹⁰¹

The conceptualisation is therefore similar to Bogert's, although the separation of liability and remedy and general judicial awareness of the position of third parties permits greater emphasis on the justice of the outcome.¹⁰² The developing approach in New Zealand is also comparable. There, a finding of unjust enrichment or unconscionability triggers the court's discretionary determination of the most appropriate remedy.¹⁰³ A constructive trust resulting from the exercise of this discretion can arise only when declared but will be backdated to the time the cause of action arose, to the detriment of creditors and those with later equitable interests.¹⁰⁴ It will be imposed, therefore, only where such detriment is warranted or such creditors and later interests are absent.¹⁰⁵

(b) The utility of the approaches

Both the “limited flexibility” and “monolithic” approaches are clearly of some merit, not least because they create a remedial framework within which specific relief can be granted without necessarily prejudicing third parties. It is recognised that the constructive trust does not simply enforce pre-existing property rights but rather operates to establish priorities between the claimant and other parties.¹⁰⁶ Thus, claimants are conferred priority over the defendant's creditors “by means of (and not by reason of) their having a

¹⁰¹ *ibid.*, at 648

¹⁰² See, *e.g. ibid.* at 651 and contrast to the treatment of the automatic retrospective approach by Oakley, who continues to view priority as an inevitable incident of the constructive trust.

¹⁰³ See *Chodar*, n.2 above, at [42]

¹⁰⁴ *Fortex*, n.2 above, at 175-177 *per* Tipping J, 179-180 *per* Henry J (referring to the creation of a “backdated proprietary interest”).

¹⁰⁵ For an illustration of the latter, see *Chodar*, n.2 above.

¹⁰⁶ See, in particular, *Barnabe v Touhey*, n.98 above, at 378-379.

proprietary interest”.¹⁰⁷ Moreover, the approach makes it difficult for a court to ignore other potential injustices created by the operation of the constructive trust because its portrayal as arising only when declared emphasises judicial responsibility for the justice of the outcome. This is more consistent with contemporary judicial attitudes in those Commonwealth jurisdictions which embrace the creative nature of the judicial function.¹⁰⁸ Indeed, without this change in legal culture it is difficult to see how the constructive trust could develop along these lines.

One might, however, question the extent to which either approach would effectively safeguard third party interests. The monolithic approach recognises that in some instances specific relief conferring general priority is warranted while in others it is warranted only to the extent third parties are not affected. In the former class of case the imposition of a constructive trust, with its retrospective effect, naturally achieves the required result. In the latter class of case the court must ensure that no third party interests are present before the constructive trust is imposed. At this point difficulties emerge because subsequent interests in or over the disputed property may not always be apparent and the solvency of the defendant may be unknown or difficult to ascertain. The defendant is the person most likely to hold the relevant information about the state of his solvency, yet least likely to make it available to the court:¹⁰⁹ it is irrelevant to his liability and there may be legitimate business reasons for his not wanting to disclose such information publicly. To avoid unintended priorities it might be suggested that it will be inappropriate to award a constructive trust without “complete information”.¹¹⁰ However, this requirement may do no more than simply provide further incentive to the defendant to withhold relevant information in order to avoid being called upon to account for his gain *in specie*.

A prospectively ordered constructive trust which takes effect from the date of judgment would not confer priority over earlier equitable proprietary interests, including those created in the interim between the time at which the claimant’s cause of action accrued and the date of judgment. However, prospective effect would not necessarily prevent the creation of unintended priorities over unsecured creditors. This may occur because a prospective

¹⁰⁷ Evans, “Property, Proprietary Remedies and Insolvency: Conceptualism or Candour?” (2000) 5 Deakin L Rev 31 at p.44.

¹⁰⁸ See, e.g. the extra-judicial writings of Kirby, *Judicial Activism* (2004); Mason, “Equity’s Role in the Twentieth Century” (1997-8) 8 KCLJ 1; McLachlin, “The Role of Judges in Modern Commonwealth Society” (1994) 110 LQR 260; McHugh, “The Law-Making Function of the Judicial Process” (1988) 62 ALJ 16; La Forest, “Some Impressions on Judging” (1986) 35 U of New Brunswick LJ 145. Justices La Forest and McLachlin in particular have been influential in the development of the constructive trust in Canada, as has Mason J in Australia: see *Lac Minerals*, n.55 above, at 44-52; *Rawluk*, n.94 above and accompanying text; *Soulos*, n.83 above, at 221-232; *Hospital Products v USSC* (1984-5) 156 CLR 41 at 110-115.

¹⁰⁹ See, in particular, *Chodar*, n.2 above, where the defendant refused to submit information regarding his solvency. Glazebrook J imposed a remedial constructive trust – something she would not have done had the defendant been insolvent – having considered evidence on the matter that was “vague and unparticularised”.

¹¹⁰ *CDIC v Principal Savings & Trust Co* (1998) 224 AR 331 at 341.

constructive trust would confer priority over general creditors in those presumably not uncommon situations in which the defendant is made insolvent by the court finding against him.

One might also question the appropriateness of forcing into the constructive trust mould specific relief that does not bind third parties. By interposing discretion between the facts creating liability and the decision to impose the constructive trust the constructive trust appears a unitary concept: certain facts create a discretionary determination of its availability and, once available, it confers on the claimant a property right. However, in those situations in which relief will be awarded only to the extent third parties are not affected, the imposition of a constructive trust appears nothing more than a misleading label to denote what is akin to a personal order to transfer property. This appears distinct from the idea of the constructive trust – a “proprietary concept” by which the claimant “is found to have an interest in property”.¹¹¹ Were the *in personam* nature of the obligation created by the order openly recognised, there would be no risk of the creation of unintended priorities. Being a personal obligation, the claimant would be left to enforce judgment in the same way as other creditors with personal claims against the insolvent. The constructive trust, as a proprietary concept, could then be left intact to refer only to those situations where it is appropriate for the claimant to take the advantages which follow from the recognition or creation of an equitable proprietary interest.

Acknowledging this may bring with it a number of benefits. First, it may contribute to the rational development of the law because it assists the schematic or taxonomic presentation of the subject, whether that taxonomy is a taxonomy of rights¹¹² or a taxonomy of remedies.¹¹³ Second, we avoid the confusion that follows from the use of one term to refer to two distinct remedies with different intended consequences. Third, it enables us to remove confusing trust nomenclature in those instances where a defining element of a trust – enforcement of interest against third parties – is absent. Fourth, the recognition that one species of “constructive trust” does not affect third parties resolves the difficulties created by the need to inquire into the liquidity of the defendant before awarding relief.

(c) The constructive trust of tomorrow: the pure remedial approach.

Some go further than the Canadian courts and suggest that greater flexibility is required to achieve justice. Most notably Professor Waters champions a

¹¹¹ These being the characteristics of the constructive trust in Canada, as stated in *Peter v Beblow*, n.88 above, at 649 *per* McLachlin J.

¹¹² As propounded by Birks, n.2 above, in which case the constructive trust and the personal order would reflect two different responses in the second measure of restitution.

¹¹³ As propounded by, *e.g.* Wright, “Wrong and Remedy: A Sticky Relationship” [2001] SJLS 300, in which case the personal order would represent a new remedy and take its place alongside the constructive trust and other potentially available remedies, to be selected according to criteria of appropriateness once liability is established. See also Waye & Wright “Trial Strategy When Selecting a Remedy From the Remedial Smorgasbord” (1998) 17 Australian Bar Rev 263.

“pure remedial constructive trust”,¹¹⁴ the court retaining maximal flexibility as to its operation and content. Its precise effects will thus vary with the circumstances of each case and the conduct of each creditor:¹¹⁵ the order may give effect to a constructive trust from the time of the unjust enrichment, the time of the court order, some interim date, or some future date; and the order may be granted on whatever terms that the court sees fit.¹¹⁶

“[The courts could] ensure that only particular third parties are affected by the *in rem* restitution. A[n] order might be given as against one defendant, but not against another, or it might be given as against both but with the content of the order varied at the discretion of the court to meet the equities of the situation, as far as each defendant is concerned. It would affect third parties not before the court by creating such priorities as the court states in the order . . . As there is no reason in my opinion why every chargee or lienholder of the assets in dispute who has acquired his charge or lien between enrichment and court order should automatically be relegated to a priority behind the claimant of the assets for unjust enrichment, so in my opinion there is no reason why every unsecured creditor of the enriched party should be so relegated.”¹¹⁷

The merits of the claim of each individual potentially affected by the *in rem* relief granted are therefore considered.

(d) The utility of pure remedial theory

The concern for greater individualised justice and fairness to third parties is laudable. In theory, the model permits the award of specific relief on terms that are always fair and just to the defendant, the claimant, and any third parties. The approach is not open to the criticism that it represents an illegitimate erosion of the *pari passu* principle where the defendant is insolvent. This principle does nothing more than affirm that which is self-evident: parties standing in positions of relative equality are to be treated equally.¹¹⁸ The pure remedial approach simply rests on a more refined analysis of relative equality. It is, however, problematic for other reasons.

First, it is difficult to imagine circumstances in which such a broad discretion to discriminate between individual creditors is necessary. And even were we to accept in principle that such circumstances might arise, it would be

¹¹⁴ Waters, “The Constructive Trust in Evolution: Substantive and Remedial” in *Equity and Contemporary Legal Developments* (Goldstein ed., 1992) 457 at p.499.

¹¹⁵ *ibid.*, p.490, n.77 and pp.499-505. See also Waters, “Trusts in the Setting of Business, Commerce and Bankruptcy” (1983) 21 *Alta L Rev* 395 at p.434.

¹¹⁶ n.114 above, at p.502.

¹¹⁷ Waters, n.114 above, at p.503. See also Waters, n.17 above, at pp.1216-1217.

¹¹⁸ McCormack, *Proprietary Claims and Insolvency* (1997), p.1; Odith, “Assets and the Treatment of Claims in Insolvency” (1992) 108 *LQR* 459 at p.463 (noting the principle “explains remarkably little”).

irrational to revise our concepts presently in the hope that they adequately accommodate that which we cannot foresee with any real precision.¹¹⁹

Second, legal forms that sit at either end of the finely shaded rules-discretion continuum are less workable in practice than in theory.¹²⁰ While a highly individualised approach theoretically optimises the chance of reaching the most appropriate outcome, the reality is somewhat different. Doctrines that seek high degrees of individualised justice increase the risk of arbitrariness and mistake, lack certainty and predictability, and require the inefficient and time-consuming application of background policies.¹²¹

Certainty, predictability and efficiency are of particular significance in the context of *in rem* relief due to the potential for detriment to third parties. The more uncertain the position of third parties who deal with the disputed property, the greater the transaction costs of property dealings and the more difficult it is for individuals to take *ex ante* measures to protect their interests and rationally determine the risks involved in their dealings. Moreover, it is a primary concern of insolvency law to minimise the cost of determining priorities: the greater those costs the less remains of the already insufficient assets to meet the claims of creditors. Efficiency rather than finely tuned justice between each affected individual therefore assumes greater significance. The pure remedial approach is antithetical to these concerns. An elaborate, time-consuming and finely tuned balancing of the merits of individuals' claims increases the role, and hence costs, of the liquidator or trustee in bankruptcy.¹²² Moreover, uncertainty and unpredictability may encourage litigiousness in the desperate creditor: "scarcity begets innovation in the hungry creditor's quest to get a little more than the next fellow".¹²³ The clearer the stated priorities the less hope there is for the desperate creditor to argue that his case is relevantly different from others in the queue, hence the less incentive to resort to litigation.

¹¹⁹ Birks forcefully reminded us of this, warning "against change which outstrips the intellect, and loses touch with the demand for stability and consistency": Birks, "Equity in the Modern Law: An Exercise in Taxonomy" (1996) 26 UWALR 1 at p.4.

¹²⁰ See, e.g. Sunstein, n.28 above; Kaplow, "Rules Versus Standards: An Economic Analysis" (1992) 42 Duke LJ 557; Schauer, *Playing by the Rules: A Philosophical Examination of Rule-Based Decision-Making in Law and in Life* (1991), Ch.7; Schneider, n.28 above.

¹²¹ Waters appears to inconsistently suggest that his discretionary model does not lie in tension with these requirements: see Waters, n.114 above, at p.504. More recently, he has emphasised the rule-building nature of the discretion and suggested a need for the principled resolution of disputes: Waters, "The Nature of the Remedial Constructive Trust" in *Frontiers of Liability* (Birks ed., 1994), Vol.2, 165 at pp.184-185.

¹²² The approach also raises the possibility of conflict between separate court orders in favour of different claimants, each creating conflicting rights against the insolvent and other specified third parties.

¹²³ *Re Omegas Group Inc*, n.49 above, at 1445.

The Institutional-Remedial Hybrid: *Muschinski v Dodds* and its Progeny.

(a) *Muschinski v Dodds* and subsequent interpretations

Similarly stimulated by a desire to avoid prejudice to third party interests, the Australian courts have developed an approach that appears to blend inconsistently both institutional and remedial characteristics of the constructive trust. Institutional elements are evidenced by the characterisation of the trust as arising at the date of the operative facts and pre-existing judicial declaration. Remedial characteristics take the form of flexibility within the operation of the constructive trust, permitting the courts to tailor the consequences to meet with the demands of the case. The conceptualisation follows the influential judgment of Deane J in *Muschinski v Dodds*.¹²⁴ Drawing on Scott,¹²⁵ Deane J noted:

“. . . notwithstanding that the constructive trust is remedial in both origin and nature, there does not need to have been a curial declaration or order before equity will recognise the prior existence of a constructive trust . . . Where an equity court would retrospectively impose a constructive trust by way of equitable remedy, its availability as such a remedy provides the basis for, and governs the content of, its imposition *inter partes* independently of any formal order declaring or enforcing it.”¹²⁶

However, later in his judgment his Honour further noted that “the constructive trust may be moulded and adjusted to give effect to the interplay of equitable principles in the circumstances of the particular case” and an order “can properly be so framed that the consequences of its imposition are operative only from the date of judgment or formal court order or from some other specified date.”¹²⁷ In this particular instance, his Honour held “[l]est the legitimate claims of third parties be adversely affected, the constructive trust should be imposed only from the date of publication of reasons for judgment”.¹²⁸ On this reasoning, the constructive trust comes into existence automatically upon the facts, prior to curial declaration, but the court retains the power to vary the date from which the trust’s consequences are operative.¹²⁹

Two interpretations of the case have followed. The broad interpretation suggests there is maximal flexibility: a constructive trust may operate from the moment of the unconscionable conduct, from the date of judgment or from any intermediate or later date, as the justice of the case demands. Deane J appears to have supported such an expansive power when he commented that the constructive trust may operate “from the date of judgment or formal

¹²⁴ (1985) 160 CLR 583.

¹²⁵ Scott, *The Law of Trusts* (1967, 3rd ed.), Vol 5, s.462.2.

¹²⁶ n.124 above, at 615.

¹²⁷ *ibid.*, at 623.

¹²⁸ *ibid.*

¹²⁹ The approach has been congratulated by many: see, *e.g.* Oakley, n.73 above, at pp.437-438 and 455.

court order or from *some other specified date*".¹³⁰ A narrow interpretation permits a more restrictive choice: the constructive trust arises prior to curial declaration, but the court has the power to determine whether its consequences take effect from the moment of the conduct giving rise to its imposition or, alternatively, from the date of judgment. This interpretation more accurately reflects the actual outcomes of the cases,¹³¹ although more recently any flexibility as to the operative date of the constructive trust has been doubted, at least in relation to common intention constructive trusts.¹³²

(b) The merits of *Muschinski* and its progeny

While both narrow and broad interpretations of *Muschinski* display a legitimate concern about the problematic effects of proprietary relief on third parties, they also raise a number of problems. To the extent that the broad interpretation represents a plea for individualised justice, it is susceptible to the same criticisms as are levelled against the pure remedial approach and contrary to the rejection by Deane J elsewhere in his judgment of abstract notions of justice and fairness when determining the appropriate relief.¹³³ Both narrow and broad approaches also risk failing to avoid unwarranted and unintended priorities due to the court's necessary dependence on accurate information about the defendant's solvency.

With its emphasis on moulding, the approach has also created confusion and uncertainty because the extent to which the court can legitimately mould and manipulate the operation of the constructive trust remains unclear. Some lower courts have established a constructive trust by recourse to existing equitable doctrine but then, rather than consider timing of imposition, have proceeded to award a quite different form of relief.¹³⁴ Other courts have further suggested that the constructive trust may be stripped of all of its consequences and no relief granted at all.¹³⁵ This is needless circumlocution.

¹³⁰ See the quote accompanying n.127 above (emphasis added). See generally Wright, *The Remedial Constructive Trust* (1998), pp.263-270; Wright, "The Statutory Trust, The Remedial Constructive Trust and Remedial Flexibility" (1999) 14 JCL 221.

¹³¹ Compare *Re Sabri* (1996) 21 Fam LR 213, *Kidner v Secretary, Department of Social Security* (1993) 31 ALD 63 and *Re Jonton Pty Ltd* [1992] 2 Qd R 105 (interest operative from time cause of action accrued) with *Re Osborne* (1989) 91 ALR 135 (common intention constructive trust not operative prior to husband's insolvency).

¹³² *Parsons v McBain* (2001) 192 ALR 772 (interest under constructive trust effective from the date the cause of action accrues, but, as a matter of priorities rather than timing of operation, may be deferred to later claims where the parties' conduct warrants it).

¹³³ n.124 above, at 615.

¹³⁴ For discussion and critical evaluation of these cases, see O'Connor, "Happy Partners or Strange Bedfellows: The Blending of Remedial and Institutional Features in the Evolving Constructive Trust" (1996) 20 MULR 735 at pp.745-751; Austin, "The Melting Down of the Remedial Trust" (1988) 11 UNSWLJ 66.

¹³⁵ *Australian National Industries v Greater Pacific Investments Pty Ltd (in liquidation) (No3)* (1992) 7 ACSR 176 at 190 *per* Cole J. The view is reminiscent of the *Restatement*, which conceptualises the constructive trust as arising automatically but being unenforceable where the legal remedy is adequate: see n.9 above, s.160, comments e & f.

Where a constructive trust “arises” but an alternative remedy is granted there is, in fact, no constructive trust. The facts simply create the grounds for the imposition of a different remedy. Talk of the constructive trust can drop out of the picture. Constructive trust terminology serves to designate a series of legal consequences attaching to particular facts.¹³⁶ If such consequences do not follow, there is no constructive trust.¹³⁷ Talk of the constructive trust can similarly drop out of the picture where its operation is apparently “suspended” and no remedy granted at all. Indeed, in such circumstances it is meaningless to talk of the claimant as having a cause of action: without a remedy there is no right.

(c) The interim equity model

In an attempt to avoid injustice to intervening third party interests, while avoiding the inconsistencies of *Muschinski*, some have argued that the constructive trust claimant neither has, nor is deemed to have, an equitable interest in the disputed property prior to the court order.¹³⁸ Rather, the potential constructive beneficiary’s interest is simply an “equity” which is transformed into an equitable interest upon curial declaration.¹³⁹ The constructive trust claimant will therefore not take priority over any equitable interest created in the interim, between the date of the facts giving rise to the constructive trust claim and the date of curial declaration, such as interests of purchasers under specifically enforceable contracts and equitable mortgages.¹⁴⁰ The potential beneficiary will, however, continue to enjoy priority over the constructive trustee’s unsecured creditors by virtue of the general principle that a trustee in bankruptcy is bound by all of the equities affecting the bankrupt.¹⁴¹

There is, however, little support for the interim equity model in the authorities of any Commonwealth jurisdiction. Indeed, there exists contrary authority on the very issue in question. Thus, in *Re Jonton Pty Ltd*¹⁴² it was held that, although a common intention constructive trust was not declared judicially until 1991, the claimant’s equitable proprietary interest existed from the time of the events giving rise to the claim. Since these occurred in

¹³⁶ Dobbs, n.20 above, at p.398.

¹³⁷ It is for this reason that Dobbs brands the analogous US orthodoxy “a fling with esoteric metaphysics”: n.20 above, s.4(3)(2), n.1. See also Palmer, *Law of Restitution* (1978) s.1.4.

¹³⁸ Most notably, Glover, “Bankruptcy and Constructive Trusts” (1991) 19 *Aus Bus L Rev* 98.

¹³⁹ *ibid.*, at pp.109-110.

¹⁴⁰ *Latec Investments Ltd v Hotel Terrigal Pty Ltd* (1965) 113 CLR 265; *Re Papaloizou* (1980) [1999] BPIR 106 (Browne-Wilkinson J); *Phillips v Phillips* (1861) 4 De GF & J 208.

¹⁴¹ *Ex parte Holthausen* (1874) LR 9 Ch App 722 at 726 *per* James LJ; *Re Clark, ex parte Beardmore* [1894] 2 QB 393 at 410 *per* Davey LJ. Similar principles apply upon corporate insolvency: see Anderson, “The Treatment of Trust Assets in English Insolvency Law” in *Commercial Aspects of Trusts and Fiduciary Obligations* (McKendrick ed., 1992) 167 at p.172.

¹⁴² [1992] 1 Qd R 105.

the 1970s the claimant took priority over the equitable interest of a mortgagee whose claim dated from the 1980s.¹⁴³

The approach also fails to develop a form of specific relief that does not bind unsecured creditors. It simply addresses the problem of priority between competing equitable interests, reversing the existing rules of priority. Is this really necessary? While there may be justification to subordinate a constructive trust claim to a later equitable interest, this will not always be so. Exceptions will be therefore required just as there is presently a perceived need for exceptions to the “first in time” rule that presumptively favours the constructive trust claimant. A sensitive and effective solution to this problem may be already within our grasp, without widespread conceptual change. We might keep the existing approach, classifying the unclaimed constructive trust interest as an equitable interest, but where required conclude that the merits are unequal so that the equity first in time (the constructive trust claimant) does not prevail.¹⁴⁴ Dealing with the matter by reference to more flexible priority rules is also likely to avoid the significant confusion that would follow the classification of a claim as an “equity”, given the protean and indeterminate nature of the term.¹⁴⁵

Undesirable and unforeseen practical consequences may also follow such wholesale change of the constructive trust claimant’s interest prior to curial declaration, thus conflicting with the principle of minimal conceptual disturbance. For example, the blanket classification of the potential constructive trust claimant’s interest as an “equity” may disturb the application of various statutory regimes in which the temporal requirement of a proprietary interest assumes significance. An example of some currency from the United States concerns the temporal requirement of an interest in property under the Racketeering Influenced Corrupt Organization Act (RICO).¹⁴⁶ This legislation allows the government to seize assets that belong to the wrongdoer at the time of any illegal action falling within the Act. Thus, in those Circuits adopting the “automatic vesting” or “automatic retrospectivity” conceptualisations the constructive trust claimant is not

¹⁴³ See also *Re Sabri* (1996) 21 Fam LR 213. Priorities between constructive trust claimant and competing interests are determined in English law in the same manner. Thus, in *Re Morgan* (1881) LR 18 Ch D 93 an executor of a deceased’s estate renewed a lease in his own name and afterwards deposited the lease as security for a loan which he applied for his own purpose. The claim of the estate took priority over that of the deposittee because the estate’s equitable interest (under a constructive trust which arose by operation of the rule in *Keech v Sandford* (1726) Sel Cas Ch 61) was first in time.

¹⁴⁴ *Heid v Reliance Finance Corporation Pty* (1983) 154 CLR 326 at 341-342 *per* Mason & Deane JJ. Such flexibility may not exist in English law, although see *MacMillan v Bishopgate Investment Management (No3)* [1995] 3 All ER 747 at 768 *per* Millett J (order of priority may be reversed in “special circumstances”); *Rice v Rice* (1854) 2 Drew 73 *per* Kindersley VC (explaining the maxim *qui prior est tempore, potior est jure* operates as a tie-breaker, to be applied only where there is no other ground for choosing between the competing claims).

¹⁴⁵ As to which, see Everton, “‘Equitable Interests’ and ‘Equities’ – In Search of a Pattern” [1976] Conv 209; Smith, *Property Law* (4th ed., 2003), pp.29-31; Skapinker, “Equitable Interests, Mere Equities, ‘Personal’ Equities and ‘Personal Equities’ – Distinctions With a Difference” (1994) 68 ALJ 593.

¹⁴⁶ 18 USC 1963.

subordinated to the government's RICO claim: the disputed property either belonged or is deemed to have belonged to the claimant, not the defendant, at the time the illegal act was committed.¹⁴⁷ However, changing the conceptualisation changes the priorities. Where the courts have accepted the view that the constructive trust, properly conceived, is a remedy that can arise and operate only when declared, it follows that the defendant and not the claimant owned the disputed asset at the time the illegal act was committed. The government is therefore able to seize the asset as an asset of the wrongdoer at the appropriate time, creating a result that is arguably contrary to the legislative intent of Congress.¹⁴⁸ Similar problems may occur closer to home, if all constructive trusts were to operate from the time of declaration only.¹⁴⁹ One might also consider the problems that would have arisen when a potential constructive trust claimant sought to assert their interest against a third party by virtue of the Land Registration Act 1925, section 70(1)(g).¹⁵⁰ If the claimant's interest was classified as a mere equity the temporal requirement that the appropriate proprietary interest be in existence at the time of the actual occupation would not be satisfied.¹⁵¹

Conclusion: Towards A New Analytical Framework

This review suggests that the various conceptual shapes into which the constructive trust has been moulded fail to work for us as effectively as they might. The automatic vesting approach persists in English law for reasons of ideology, not utility. It avoids consideration of crucially important normative concerns and potentially perpetuates unjust and largely unintended judicially created priorities. Its operation and interplay with related doctrine lacks logic and often depends on reasoning that is, at best, grossly artificial. Oakley's automatically retrospective conceptualisation avoids some of these difficulties but similarly portrays priority over third parties as an inevitable and natural consequence of the constructive trust, reflecting the traditional reluctance of English judges and jurists to engage openly in critical normative issues affecting priorities.

Other approaches, developed against the background of more progressive judicial cultures, have been motivated by the desire to address openly the very issue that the institutional constructive trust suppresses: the justice of the constructive trust claim against innocent third parties. Developments in Canada are perhaps the most promising. There, it is recognised that in some

¹⁴⁷ See *US v Schwimmer* 968 F 2d 1570 at 1582 (2d Cir 1992) and *US v Lavin* 942 F 2d 177 at 187 (3d Cir 1991).

¹⁴⁸ *US v BCCI Holdings* 46 F 3d 1185 (DC Cir 1995).

¹⁴⁹ A subsisting equitable interest may be of importance in England for taking otherwise relevant property outside the operation of similar confiscation regimes, because such regimes will respect third party rights in property subsisting at the time of confiscation order: see, e.g. *HM Customs & Excise Commissioners and Long v A* [2002] EWHC 611 at [171], affirmed [2002] EWCA Civ 1309 at [23]-[24], [50]-[54], discussing the Drug Trafficking Act 1994, s.31(4). See now Proceeds of Crime Act 2002, s.69(3).

¹⁵⁰ See now Land Registration Act 2002, Sch.3, para.2.

¹⁵¹ See now Land Registration Act 2002, s.116, by which a mere equity is capable of binding successors in title. Note, however, that the language of "mere equity" may not be taken literally: Smith, n.145 above, at p.31.

cases the constructive trust binds third parties legitimately; in others it does justice between claimant and defendant only to the extent that innocent third parties are not prejudiced. The problem is that the latter cases continue to be dealt with via the language of constructive trust. This makes “constructive trust” do too much and obscures the individuality of what is in substance a different remedy with different intended consequences. It also fails to avoid the risk of unintended and unwarranted priorities because it places such a heavy reliance on the availability of information about the defendant’s solvency. There is reason to believe that the quality of such information will be often poor, leaving a court to second guess the appropriateness of the imposition of a constructive trust. The hybrid approaches developed in Australia post-*Muschinski* are problematic for similar reasons.

These pitfalls are avoided, and justice and clarity enhanced, by the recognition of two conceptually distinct remedies: the constructive trust and the personal order to transfer specific property. The constructive trust is more appropriately conceptualised as a judicially imposed interest in property operating retrospectively from the time the claimant’s cause of action accrued. It should be given a reduced sphere of operation and imposed only where there is reason to grant the claimant the significant benefits that follow from the recognition or creation of equitable property rights, particularly priority over creditors and supervening interests. By this approach, judicial responsibility for the outcome is emphasised. Moreover, the gross artifice and logical deficiencies of the automatic vesting approach are avoided.

The purely personal order to transfer specific property would exist alongside the constructive trust.¹⁵² It would operate in those cases previously dealt with by the constructive trust in which specific relief is desirable but priority is not. It would, for example, provide a more satisfactory way forward in disgorgement cases. In these cases recovery is premised on the desire to deter wrongful conduct rather than correct a material loss.¹⁵³ What is

¹⁵² Professor Goode alludes to the need for specific relief that does not prejudice creditors. However, he limits the operation of the device to “deemed agency gains”, would continue to recognise an automatic vesting constructive trust, and is ambivalent as to whether such a remedial order should confer on the claimant a prospective proprietary interest or a purely personal right to the transfer of specific property: see Goode, “Property and Unjust Enrichment” in *Essays on the Law of Restitution* (Burrows ed., 1991) 215; Goode, “The Recovery of a Director’s Improper Gains: Proprietary Remedies for Infringement of Non-Proprietary Rights” in *Commercial Aspects of Trusts and Fiduciary Obligations* (McKendrick ed., 1992) 137 at pp.146-148.

¹⁵³ The most obvious cases are those in which a fiduciary receives a secret profit, such as a bribe or profitable opportunity, from a third party in breach of his duty of loyalty: see, e.g. *Attorney General for Hong Kong v Reid* [1994] 1 AC 324; *Boardman v Phipps* [1967] 2 AC 46. This is not to say, however, that all bribe and opportunity cases fall out with the category of correction of material loss. It is to say simply that some do. I do not therefore include as instances of disgorgement those bribe and opportunity cases in which the principal can demonstrate a loss corresponding to the defendant’s gain by direct or interceptive subtraction: see, e.g. *Daraydan Holdings Ltd v Solland* [2004] EWHC 622 (Ch) (bribe subtracted from claimant) and *Cook v Deeks* [1916] 1 AC 554 (interceptive subtraction of corporate opportunity).

therefore important is not that the claimant receives the benefit but that the defendant does not. The claimant is simply an “accidental beneficiary of a rule of public policy”.¹⁵⁴ Consequently, it is difficult to find justification for elevating the claimant to the status of secured creditor.¹⁵⁵ A purely personal order to transfer specific property would enhance the law’s ability to effect full disgorgement where the gain is non-fungible without risking the creation of unintended and unwarranted priorities. Others argue for a similar result but by reference to the distorting language of discretion and the remedial constructive trust. They argue that a court should possess discretion to impose a constructive trust, but that discretion should not be exercised positively where the defendant is insolvent or supervening third party interests are present.¹⁵⁶ This does not sound like a proprietary remedy, nor does it sound like discretion. It is a call for the imposition or recognition of an *in personam* obligation to disgorge a gain *in specie*. The notion of a purely personal order to transfer specific property more faithfully represents what is being sought. It also affords a way of avoiding unintended priorities and the expense and difficulty of an inquiry into the defendant’s solvency.

This, of course, leaves much unsaid about any reclassification. Which cases should remain within the scope of the constructive trust? Which are more appropriately dealt with by the personal order to transfer specific property? And are there other areas to which the personal order, free from priority problems, could be extended? These questions may prove difficult and generate controversy, but they raise issues that are there already and which will not go away. The benefit of the framework for analysis proposed here is that it clarifies the choices to be made and illuminates what is at stake in making them.

¹⁵⁴ *McCarey v Associated Newspapers Ltd (No.2)* [1965] 2 QB 86 at 107 per Diplock LJ.

¹⁵⁵ See, e.g. Sherwin, n.18 above, at pp.329-340; Paciocco, “The Remedial Constructive Trust: A Principled Basis for Priority Over Creditors” (1989) 68 Can Bar Rev 315 at pp.349-350. Indeed, given that the claimant is receiving a windfall, it is difficult to find justification for granting a disgorgement claim even equal status to other unsecured claims: see, e.g. Jaffey, “Disgorgement and Confiscation” [1996] RLR 92 at p.97.

¹⁵⁶ See, e.g. Cope, “Ownership, Obligation, Bribes and the Constructive Trust” in *Equity Issues and Trends* (Cope ed., 1995) 91.

'PROTESTANT' POLITICAL THEORY AND THE SIGNIFICANCE OF RIGHTS

*Sean Coyle, Reader in Jurisprudence, University College London**

Modern political thought might be presented as a tension between two competing ideals: first, the notion of a community as a form of moral association which establishes and protects individuals' interests through representative politics and collective mechanisms of enforcement (which I shall refer to as 'idealism'); and, second, the idea of society as a dog-eat-dog world in which the private pursuit of wealth and happiness is constrained by rules and standards operating for the benefit of all (which I shall call 'positivism'). The notion of a right is central to both of these ideals, and it is plausible to see theories of rights as offering differing understandings of the significance of the modern polity's attachment to liberal values. The notion of a 'right' is of enormous significance for the liberal polity. Modern jurisprudential thought distinguishes two rival theories of rights, known as the Will theory and the Interest theory. My aim in this essay is to defend the Will theory of rights. Jurisprudential theories of rights are often presented as analytical attempts to clarify the essential nature of legal rights. The present essay represents a departure from this tradition: it is my aim to suggest that, far from constituting a dispute about the proper analysis of a concept, the Will and Interest theories represent fundamental yet conflicting ideologies which each throw differing light on our form of political association.

Two Traditions of Political Reflection

“[T]here is one kind of good that is so called in an absolute sense, and there is another that is good from the standpoint of a particular individual. Indeed, to borrow Aristotle's admirable explanation, 'Whatever each person's understanding has ruled for him regarding a given matter, that to him is good.' For God created Man . . . 'free and *sui iuris*', so that the actions of each individual and the use of his possessions were made subject not to another's will, but to his own . . . For what is that well-known concept 'natural liberty' other the power of an individual to act in accordance with his own will?”¹

This passage of early modern political theory, and others like it, represents a crossroads in political thought. Having shaken off medieval assumptions concerning the association of the divine with the mundane, the idea began to take shape in the writings of Grotius and his contemporaries, that a form of life might be related to general ideas of natural law through the development of customary principles and forms of action which arise in response to

* I would like to thank Matthew Kramer, George Letsas and Amanda Perreau-Saussine for their helpful comments and suggestions, as well as participants at a seminar in Queen's University Belfast where some of the arguments herein were presented.

¹ Grotius, *De Iure Praedae Commentarius* [1604] (Oxford, Clarendon Press 1950), 18.

particular problems of communal living. Reflection on our moral life, then, is not simply a matter of reporting the content of such customs and forms of action, but it is nevertheless rooted in those forms of action. The willingness of Grotius and his contemporaries to recognise distinct forms of the good constitutes a form of moral Protestantism the effects of which are with us still. For, on the one hand, we are accustomed to thinking of our political arrangements and social practices as being open to criticism in the light of moral standards which, in some sense, transcend those practices; and, on the other hand, we tend to view such critical standards as a matter of irresolvable disagreement and subjective understanding. These thoughts present us with conflicting views about the nature of moral insight: morality might be understood as a set of standards and principles which exist apart from the historically extended practices to which they apply, or as something fundamentally rooted in reflection on the here and now. The political theory of the modern day might then be thought of as the attempt to reconcile these divergent views about the source of moral insights with the thought that morality embodies standards and interests which make sense only within a system of collective practices.

The history of political thought has tended to emphasise two distinctive responses to the problems posed by moral Protestantism. One way of bringing the complexities of politics within the scope of general moral principles is to view politics as the earthly realm in which eternal values are applied and enforced. We can refer to this understanding of political life as 'moral idealism', or simply, 'idealism'. Political values, on this view, are the concrete manifestation of general moral demands, rather than a distinct kind of demand brought about by conditions of scarcity, need and inexhaustible desire. Insofar as morality concerns interests that are held and asserted within the context of shared social practices, the political life of the polity must be geared towards the resolution and protection of clashing interests according to standards of moral rightness which transcend the understandings which individuals bring to the identification of their interests and claims. Alternatively, the political realm might be represented as a domain of clashing interests which can be regimented in various possible ways: moral standards function not only within a pure realm of abstract thought, but also within the real world of social tensions and conflicting demands. Politics is then a matter of constructing and testing those possibilities. We can refer to this second view as 'positivism'.² Rather than constituting a continuous intellectual realm, morality and politics can be construed as defining separate yet overlapping intellectual domains, in which political values represent generalised responses to the variable conditions of collective living, offering makeshift solutions to the problem of continually shifting desires and perpetually clashing interests. Customary norms and forms of redress may then evolve in response to shared expectations, but we might regard those forms of redress as being related to general moral norms only broadly and indirectly.

² I do not, of course, mean to refer to the view that there are no 'necessary connections' between law and morality, or to the view that law consists only in rules which have been deliberately laid down by some recognised authority. Positivism in this sense nevertheless has obvious and important connections with more familiar views which go by that name.

The notion of a 'right' is central to both approaches. The implications of these connections may, however, be traced out in different ways. For, on the one hand, it is possible to conceive of rights as establishing and protecting individual interests (the 'Interest theory'), and on the other, rights might be viewed as securing individual freedom from collective control, through the establishment of protected choices (the 'Will theory'). Initially, at least, we might be led to assume that such distinctions are of no significance to the distinction between positivist and idealist views of the polity.

The idealist views the legal order as outlining principles of justice which apply *mutatis mutandis* to all citizens, who stand in social relationships requiring systematic expression. We might view the idealist notion of rights as one in terms of which an individual's rights are determined on the basis of general principles which underpin the polity's aggregative and distributive goals. We will then view rights as embodying interests defined on the basis of judgments about the common good. Alternatively, idealists might conceive of individual rights as mapping out conjointly possible domains of autonomy in which individuals are free to pursue their own interests in their own way. Judgments as to the extent of individual rights would then be a matter of determining which exercises of choice are compatible with similar exercises of choice on the part of others on the basis of principles of equal freedom: rights, on this view, constitute protected choices rather than legally established interests.

Positivists, by contrast, view the doctrinal categories of the law in a more limited way, as outlining a series of rules and principles which achieve a reasonable balance between competing interests and which establish powers through which individuals can make normatively significant choices about their relationships with other people. The focus of the positivist account is therefore not upon the articulation of general principles which apply as political standards to all, but on the delineation of rules which provide a framework for social interaction, and on the interstices between the rules within which individuals may pursue interests and exercise choices.

Despite the possibility of such connections between positivism, idealism and the notion of a right, there are good reasons for disputing the suggestion that the distinction between protected interests and protected choices has no significance for the political ideals represented by positivism and idealism. Indeed, my arguments will seek to establish that the notion of protected interests is central to the possibility of liberal forms of idealism, and that positivism is best understood in association with the protection of individual choice. One reason why the idea of a 'right' might be thought to have no intrinsic significance for debates about the nature of the polity is that the jurisprudential analysis of rights is thought to highlight different ways in which our conceptual intuitions about the ordinary use of the word 'right' might be regimented: by, for example, denying that children can correctly be said to possess rights, or by suggesting that rights reflect public standards of conduct (within the criminal law, say) as well as private choices. Because no political issues are predetermined by such definitional preferences, it is thought, the notion of a right embodies a set of conceptual ideas which

remain separate from broader political ideals.³ This, I shall argue, is a mistake.

Suggestive Connections

The Will and Interest theories represent opposing conceptions of the social framework within which political projects are formulated and carried out. For they offer radically different interpretations of the *purpose* of our form of political association, and of law's role in shaping that form of association. If the life of the modern state embodies a 'Protestant' view of political theory, then the notion of a right is central to our understanding of that life.⁴ Yet neither Protestantism, nor the idea of rights, is inevitable wherever law emerges as a distinctive force in the political realm. The laws of ancient Greece, for example, were geared towards the pursuit of excellence and of rules bringing about conditions which make excellent, valuable lives possible, rather than to the establishment of rules promoting sectional interests or autonomy. Even within medieval common law scholarship, the law was conceived in terms of the reasoned settlement of wrongs, without any parallel notion of individual *rights* as we have come to understand the term.

The history of the idea of rights is accordingly an important and instructive one as far as the dynamics and tensions of modern political thought are concerned, and of attempts to understand the place of law within its peculiar currents. The language of rights comes to us from Roman law, and its origins lie in the word *ius*.⁵ The dominant usage among the Roman jurists treated *ius* as signifying something objectively right or just, and in this way it functioned as a synonym for 'law' as long as the dominant legal treatises

³ See e.g. MH Kramer, 'Rights Without Trimmings' in Kramer, Simmonds & Steiner, *A Debate Over Rights: Philosophical Enquiries* (Oxford, Clarendon Press 1998). The attribution or denial of rights to children is thus a formal consequence of the definition of a right as a protected interest or choice, and leaves undisturbed the question of the extent to which the law ought to protect the interests or choices of children. Similarly, the argument about whether the criminal law confers individual rights involves no disagreement about which forms of behaviour should be classified as crimes (or of the consequences of such classifications), but rather reflects a disagreement over the conceptual distinction between public and private law.

⁴ The combined centrality and irreconcilability of the Will and Interest theories might help to explain the tendency of some jurisprudential writers to engage in attempts to 'move beyond' those theories, or to suggest that the Will and Interest theories are not jointly exhaustive of the domain of rights. (See for example R Cruft, 'Rights: Beyond the Interest Theory and the Will Theory?' 23 *Law and Philosophy* (2004) 347-97, and W Lucy, 'Rights, Values and Controversy' V *Canadian J of Law and Jurisprudence* (1992) 195-213.) When viewed as a central tension in the political life of the state, however, such attempts to reconcile the Will and Interest theories come to resemble contrived and superficial exercises in presenting an inescapable political choice as a definitional controversy capable of being sidelined as an interesting though largely fruitless analytical problem.

⁵ More particularly, the origins of the notion of a 'right' lie in the complex and shifting relationship which existed between the concepts of *ius*, *dominium* and *res*. I do not propose to examine this relationship in any detail, but a brief yet informative account can be found in R Tuck, *Natural Rights Theories: Their Origin and Development* (Cambridge, Cambridge University Press 1979) ch.1.

were composed in Latin. At the same time, the term was essential to claims about the way two disputants should behave towards one another, and was thus confined to the description of private, bilateral relationships.⁶ Both Grotius and his English counterpart, Thomas Rutherford, were conscious of these various dimensions to the word 'right': 'right' was said to signify, first and foremost, what is right and just; but it can also carry the sense of 'law', and it could also denote an individual moral faculty or power.⁷ This ambiguity in the notion of a 'right', which could function as an objective claim *about* justice or a subjective appeal *for* justice, can be seen as the defining problem for modern political theory. For once we are in possession of the distinction between objective and subjective *ius*, the notion of objective right can be conceived as a fixed standard against which the inequalities and imperfections of ordinary life can be measured. Hence it becomes easy to imagine that the ordinary, subjective claims individuals make in relation to their rights are ultimately governed by an ideal moral order of rightness. We might then hope that the constantly shifting pattern of uncoordinated and competing claims can be rationalised in a way which reflects that ideal order: then we will have simultaneously reconciled those clashing interests and delineated the scope of individuals' *legitimate* interests, or rights.

But for the positivist, who dismisses the possibility of reconciling the notions of objective and subjective right, the political life of the state will come to be regarded as a matter of achieving a reasonable and stable balancing of those interests. Rights will then take on a different role: for if no stable perspective exists against which the irregularities and conflicts of everyday life can be evaluated and resolved, individual rights will come to play a central role in the delineation of protected spheres of interest and choice distinct from the aggregative and distributive policies of the state. The laws of the polity will accordingly resemble less a body of principles aiming at the establishment and protection of equal rights, and more a framework of social rules for dealing with the effects of manifest inequalities. To that extent, the law can be seen as displaying a concern with the needs and wants of individuals, and as securing a framework of stable expectations in which those needs and wants can be pursued and fulfilled.

The notions of *ius*, *dominium* and *res* had been the subject of technical debates by the Roman jurists for centuries. But while such arguments were a vital part of the theoretical underpinnings of Roman law, they did not figure greatly in the day-to-day applications of that law. Ideas of right appeared in the guise of general juridical assumptions rather than as weapons to be wielded in the course of Roman litigation. The circumstances of early modern Europe were nevertheless to propel those ideas and assumptions onto the centre stage of political and legal thought where, in effect, they have remained.

The collapse of medieval beliefs which linked governance and authority with the divine order left in its wake a series of assumptions about human equality. Natural lawyers such as Grotius had shown that long-established

⁶ *ibid.*, 8.

⁷ See Grotius, *De Iure Belli ac Pacis* [1625] I.1.iii-ix. Rutherford echoes Grotius in his *Institutes of Natural Law* [1754] Book 1, ch.2, 25-33.

social roles were not the reflection of some higher natural order, but rather one of a number of social forms which might be 'patterned after nature's plan'.⁸ Forms of political authority were thus to be construed as productions of the human will which attempted to give expression to mankind's essentially social nature: in the absence of such social forms, individuals confronted one another as fundamental equals. A new conception of the significance and purpose of political society emerged, not as a natural means of upholding divinely ordained values, but as a form of association in which every person's interests could be peacefully coordinated and protected.

The natural lawyers and their positivist critics nevertheless construed the significance of this fresh insight into the political realm in very different ways. On the one hand, the natural lawyers viewed the purpose of political reflection as involving the working out of the ideal form of society, in which man's natural equality is respected and enhanced. The legal life of the polity consisted, for those jurists, in tracing out the structure of an ideal body of rights in terms of which individual interests are articulated, and the legitimate bounds of individuals' entitlements discovered. Individual rights were therefore conceived as essentially public standards, a matter for collective determination in the light of a shared conception of justice rather than for private choice. For positivists, on the other hand, the evaporation of universal moral norms signalled a right to determine the course of one's life in one's own way. For the joint pursuit of individual conceptions of the worthwhile life to emerge as a genuine possibility, however, each individual must be free in various respects from collective forms of control so that autonomous choices can be meaningfully formulated and exercised. On this view, the legal order is best represented not as a systematic body of rights rooted in some shared conception of the good, but rather as a body of rules acting as anchor-points for autonomous choice in a sea of ever-shifting political assumptions.

We can, in this way, view the idea of a 'right' as suggesting differing ways of dealing with the consequences of Protestant political thought. Those positivists who saw rights as defining areas of individual autonomy shielded from collective control developed theories of law which attempted to divorce juridical conceptions and arguments from wider matters of moral and political concern. The development of detailed doctrinal rules and principles in isolation from collective political introspection was seen as central to the possibility of individual rights. For it is only in the context of such stable practices that individuals can form reliable expectations which make possible the possession of meaningful control over aspects of their lives. The dependence of private law upon general recognition and collective enforcement meant that the realm of private entitlements could never be *fully* distinct from the aspirations and goals of the public sphere, but the creation of a comparatively stable set of rules which are relatively immune from direct pressure from shifting political attitudes was nevertheless felt to be the only meaningful way in which a society could give effect to the liberal ideal. But where rights are thought of as being rooted in interests rather than the will, the integrity of rights ceases to be a matter of the comparative stability and immunity of an individual's entitlements from revision in the light of

⁸ Grotius, above n.1, 229.

principles of general political concern, and becomes a question of the integrity and consistency of a body of principles inherently and pervasively shaped by publicly determined standards and shared conceptions of the good. The liberal nature of society, then, is reflected in the view that questions of individual interest are a matter for joint interpretation and concern.

The positivist and idealist versions of Protestant political thought can be seen to reflect differing views of the liberal polity's commitment to liberal values. The foregoing reflections suggest one way in which these connections might be understood. Positivism is not *inevitably* associated with a focus on individual autonomy over a concern with individual interests, for we might view the task of law as that of achieving some reasonable balance between essentially competing interests. It is this thought which has led to the belief that positivism is equally compatible with the Interest theory as with the Will theory. Yet there are reasons for doubting the value of such connections: the interests which are established and protected by legal rules can be viewed as establishing *rights* only if we are willing to downplay the idea that the society's legal arrangements take the form of relatively fixed and stable rules. For where rights are defined in terms of protected interests, such rights are not merely the fixed and stable products of the enactment of particular legal rules, but are central to the interpretation of a body of laws as a whole. Each rule will then be read as a partial attempt to trace out the relative boundaries of competing entitlements, so that rights and interests stand apart from and above the posited rules as such. The form of moral reasoning at work in such a conception is one in which rules are reduced to the status of mere guides to the content of rights and interests, which must be fleshed out according to the political or moral values which the rules serve; or else legal rules must be viewed as highly incomplete expressions of legal rights which leave much to be determined on the basis of extra-legal, moral and political ideas.⁹ Even where positivism is understood primarily as a set of beliefs about law's constitutive role in constructing the social world, its associations with the Interest theory might thus be seen as more superficial than its connections with the realisation and protection of liberty and choice.

Idealism is characterised by the belief that rights, rather than rules, constitute the basic elements in legal thought. The development of the Will theory suggested another way in which such thoughts could be given substance: rights might be seen as determining jointly possible domains of choice sustained by principles of autonomy which apply to all. Abstract principles of autonomy accordingly define the maximal set of freedoms which can exist conjointly with similar freedoms possessed by others. Since certain exercises of freedom will restrict (or bring to an end) the freedom of others, legal rules restricting individual freedom (such as the freedom to commit violent acts against the person) are justified insofar as they are necessary for

⁹ The latter approach is the one which has in fact commanded the attention of many recent positivist thinkers (thinkers as diverse as Jeremy Bentham, Neil MacCormick, Joseph Raz and Matthew Kramer). As my remarks no doubt imply, I remain uninterested in versions of positivism which consist simply in the proposition that law and morality are conceptually separate: I continue to believe that 'positivism', in the interesting sense of a body of relatively stable rules which are interpreted in relative isolation from wider moral and political currents of thought, does not sit well with forms of collective reflection.

the realisation and preservation of freedom overall.¹⁰ This notion of a system of equal freedoms is, of course, a juridical fantasy: in a world of scarce resources and finite possibilities, my choices inherently impinge upon those of others, and a system of rules which seeks to impose uniform conditions of equal choice on a social world governed by uncoordinated wills and constantly shifting opportunities will be forced to seek for such conditions at ever increasing levels of abstraction.¹¹ The emptiness of the idea of juridical equality rooted in the will has the consequence that only forms of idealism which represent rights as interests are capable of providing meaningful insights into Protestant political thought.

The notion of a 'right' is thus not a formal one which is devoid of important political commitments, but rather one which presupposes differing conceptions of our form of political association. The remainder of this essay will describe and chart some of these commitments.

The Interest Theory and Liberal Equality

Both the positivist approach and the liberal idealist approach run into problems. Where positivism is committed to the Will theory, it might be seen as condoning forms of radical conservatism which leave the realisation of important individual interests to market forces and the vagaries of representative political institutions. By using law to preserve and extend the private realm, does not the positivist threaten important individual interests and render minority oppression at the hands of the majority much more likely? Idealism, on the other hand, by construing rights directly in terms of interests, connects questions of social and political justice much more firmly and directly with the adjudicative process. Unlike the wider political process, in which each person's voice is drowned out by an ocean of similar voices, the adjudicative process is one in which each person's claims can be heard and evaluated specifically and directly. But the same conditions which make idealism an attractive philosophy of law also undermine the value and purpose of rights as the principal means through which individuals can assert and defend their interests. By threatening the integrity of private law as a body of principles distinct from general political debate, idealism casts into doubt the reality of blocks of private life marked off from public scrutiny and regulation.

The traditional conception of adjudication is one in which parties to a dispute submit pleas to the court and present reasoned arguments and proofs for a decision in their favour.¹² It is this dimension of adjudication that lies at the centre of idealistic assumptions about the judicial process as a special and

¹⁰ The Kantian Will theory, *e.g.* was of this kind (see F Rosen, *Kant's Theory of Justice* (Ithaca, Cornell University Press 1993). For positivism's relationship with the Interest theory, see above n.8.

¹¹ See Rosen, above n.10, and N Simmonds, 'Rights at the Cutting Edge' in Kramer, Simmonds & Steiner, *A Debate Over Rights: Philosophical Enquiries* (Oxford, Clarendon Press 1998). For a recent attempt to rehabilitate the Kantian view, see H Steiner, *An Essay on Rights* (Oxford, Blackwell 1994), and a partial retraction in Steiner, 'Working Rights' in *A Debate Over Rights, op. cit.*

¹² L. Fuller, 'The Forms and Limits of Adjudication' 92 *Harvard L Rev* (1978) 353, at 364. See also O. Fiss, 'The Forms of Justice' 93 *Harvard L Rev* (1979) 1.

essential form of participation in legal decision-making through which individuals can directly articulate and protect their fundamental interests. In his impressive article 'On the Form and Limits of Adjudication', Lon Fuller presents an argument which we can see as constituting a devastating problem for idealism and the Interest theory. Fuller develops the idea of a 'polycentric' problem: a problem in which the solution depends upon the resolution of tensions within a complex web of relationships where '[a] pull on one strand will distribute tensions after a complicated pattern throughout the web as a whole.'¹³ Problems of this kind bear certain similarities to the behaviour of a spider web:

"Doubling the original pull will, in all likelihood, not simply double each of the resulting tensions but will rather create a different complicated pattern of tensions. This would certainly occur, for example, if the doubled pull caused one or more of the weaker strands to snap. This is a 'polycentric' situation because it is 'many-centred' – each crossing of strands is a distinct centre for distributing tensions."¹⁴

Where rights are thought of as protecting interests, such problems are brought to the fore: for the parties to a case are regarded as asserting claims which depend not just upon the history and circumstances of the dispute between them, but upon broad and general principles which apply to all. An individual's rights thus appear in the guise of interests wielded and asserted in the context of complex patterns of similar and competing interests, delineated not simply by reference to established rules but according to broad conceptions of equality and the good. The 'legal' dimension to adjudication is then preserved by the representation of such overtly political concerns as elements in the construction of 'juridical equality'. Yet the presence of such political elements in juridical reasoning comprehensively undermines the driving insight that the judicial process offers individuals a direct means of defending their interests against outside intrusion. By requiring rights to be delineated in a broader context of competing political assumptions, rights function effectively only insofar as they conform to collective notions of what each individual's legitimate interests are. In such circumstances, the idea that adjudication furnishes individual litigants with meaningful control over their own interests is an empty form indeed.

Fuller believed that polycentric problems constitute the limits of effective adjudication.¹⁵ For in the context of clashing political values and ideals, the ability of the litigant to offer reasoned arguments and proofs is fatally undermined: the plaintiff cannot *prove* the existence of an entitlement, because such proofs presuppose the existence of a stable structure of rules and doctrines against which competing claims can be evaluated and tested. The possibility of offering reasoned arguments in favour of a given standpoint is similarly undercut, as the process of reasoning is one no longer

¹³ *ibid.*, 395.

¹⁴ *ibid.*

¹⁵ Fuller's reasons differ rather considerably from mine, but his arguments might be seen as addressing parallel and perhaps wider concerns. There is a great deal in Fuller's careful treatment of polycentric problems which merits consideration. See Fuller, *ibid.*, 394.

rooted in a context of stable assumptions and accepted starting-points upon which the litigants' 'reasons' can focus. Rather than offering reasoned interpretations of established doctrines and principles, the party wishing to assert that a right has been infringed must orientate his claims within a context of political and moral thought where disagreement attaches not just to the interpretation of rules, but also to the question of what the permissible starting-points of the reasoning process are. Potentially intractable political and ethical dilemmas would then have been converted into legal questions which lie at the very heart of our doctrinal understandings.¹⁶

The Interest theory has a particularly powerful association with idealist modes of thought. For in making individual interests the focus of an account of rights, the Interest theorist is forced into thinking of rights as concerning not the autonomous decisions of individuals who make choices in the light of their own conceptions of their interests, but instead the general social context in which those choices are made. Since individuals disagree about their rights and interests, and since those disagreements arise from decisions made in the context of social conditions of scarcity, competition, cooperation, economic variability and ever-changing desires, rights come to be seen as a matter for collective decision as part of a political process in which other social goods are at stake. Interests become a matter of what individuals *would* recognise if they were free of the imperfections and inequalities (both advantageous and disadvantageous) which pervade their lives. The Interest theorist thus simultaneously places rights at the centre of political thought, and undermines the idea of a right as a stable and lasting form of protection whose exercise lies predominantly within the control of the right-holder rather than 'another's will'.¹⁷

Liberal idealism of this kind can plausibly be viewed as one way in which to work out the implications of a Protestant conception of political theory. The political thinking of the post-medieval world is shaped by the notion of individuals as both a focus for moral and political concern, and as sources of moral insight and reflection in relation to the good life. The erosion of religious belief as a source of shared insights and understandings, as well as the decline of the feudal order as a framework for the regulation of narrow and inherited social roles, forced the idea of conflicting individual claims into the limelight of political thought. The realm of politics was thus connected with morality in a much more complex way than medieval thought supposed: natural law came to be regarded not as a set of unchanging moral prescriptions identifying the good independently of concrete social arrangements, but as a body of broad principles capable of grounding agreement on general matters such as the desirability of social coexistence, whatever form such arrangements might take. More concrete moral insights arose, in Hume's words, 'from the circumstances and necessities of

¹⁶ One recent advocate of human rights-inspired idealism is openly circumspect about the possibility of proving rights claims: see F. Klug, *Values For a Godless Age: The Story of the United Kingdom's New Bill of Rights* (London, Penguin 2000), 18. Such claims take place in the context of a 'debate without end' where understandings of rights shift in the light of changing moral, philosophical and political ideals.

¹⁷ Grotius, above n.1.

mankind’ about which different individuals might disagree.¹⁸ Thus, while goodness could not depend solely on the content of local customs and practices, it was no longer straightforwardly possible to hold up social arrangements to the blinding light of eternal and transcendent standards: moral ideas, though constituting standards of criticism for concrete social arrangements, could not be understood wholly apart from the history of social and political institutions. Idealism, then, reflected Hume’s belief that ‘Tho’ rules of justice be artificial, they are not arbitrary.’¹⁹

Liberal idealism should be read as an attempt to tackle the consequences of moral Protestantism by displaying a concern with the general framework within which different individuals formulate and argue about their interests. The basic insight that plural visions of social and political good might coexist in a single social setting is thus recast as a concern with equality among individuals who realise that ‘the issues of principle affecting them – the people – should be settled, ultimately, by them and only them on a basis that paid tribute to their fundamental equality.’²⁰ Rather than mapping out areas within which the exercise of distinctive visions is possible, rights come to be understood as the general, public standards which result from collective deliberation over differing moral visions:

“Integrity expands and deepens the role individual citizens can play in developing the public standards of their community because it requires them to treat relations among themselves as characteristically, not just spasmodically, governed by these standards . . . Political obligation is then not just a matter of obeying the discrete political decisions of the community one by one, as political philosophers usually represent it. It becomes a more Protestant idea: fidelity to a scheme of principle each citizen has a responsibility to identify, ultimately for himself, as his community’s scheme.”²¹

According to liberal idealism, then, rights mark out important interests in which every individual has a stake, and which result from collective reflection into the nature and characteristics of individual wellbeing as a

¹⁸ Hume, *A Treatise of Human Nature*, P.H. Nidditch trans. (New York, Oxford University Press 1978) 477.

¹⁹ *ibid.*, 484. ‘Artificial’, in Hume’s sense, referred to the adventitious conditions in which justice emerges, rather than to any sense of justice as determined by purely positive stipulation.

²⁰ Waldron, *Law and Disagreement* (Oxford, Clarendon Press 1999) 249.

²¹ R. Dworkin, *Law’s Empire* (London, Fontana Press 1986) 190. Dworkin has elsewhere elaborated on this idea: ‘[I]n my view, integrity is itself a dimension of equal respect. Indeed, in a community divided in moral and political judgment and instinct, it is a peculiarly important dimension of equal respect. We cannot expect of our fellow citizens, either as individuals or collectively in politics, that they treat us as we think justice on the right conception requires. But we can demand of ourselves and of them that we treat each other evenhandedly, that is, that we deny no one the respect we accord others according to our own convictions about what equal respect means. Law as integrity assumes that we can make sense of and enforce that demand collectively, in politics and adjudication, as well as individually. . . .’ ‘Integrity, Equality and Respect: A Reply’, *Revue Internationale de Philosophie* (2005) 40.

source of interests. One might well wonder to what extent such a conception of political liberalism remains moored in the values of liberal pluralism from which the liberal tradition emerged.²² Leaving that consideration on one side, however, we may worry independently about the implications of such a view for understandings of the nature of legal reasoning. For the view of rights as interests defined in the light of some general notion of 'liberal equality' seems very far removed from the history and traditions of common law scholarship through which law's representation as a form of moral association is conceived and expressed.

The liberal equality approach compels us to look upon the law as a coherent expression of justice governed by categorical principles. The possibility of presenting legal rules and decisions as instances of more general principles and categories appears as no accidental consequence of academic scholarship, but as the result of a coherent moral vision running throughout the law. Law can be represented as a form of moral association, it is thought, precisely *because* it gives expression to a substantive moral ideal. The liberal credentials of the theory are preserved through the notion that the 'substance' of these moral ideas must be connected with jurisprudential understandings of the established rules and doctrines of the legal order, and that each person may 'fit' those substantive ideas to the rules and decisions in different ways. The reality of such connections is open to doubt: the idea of equality signifies nothing unless connected with specific situations which may be compared as equal or unequal in various respects. But the concept of equality itself can do very little to clarify which peculiar features of a case are 'relevant', so that there are potentially limitless ways in which various cases can be presented as embodying specific equalities or inequalities. Any theory which locates the law's moral nature in the idea of a substantive moral ideal is thus forced to justify general and inevitably contentious understandings of the nature of equality.

Faced with such difficulties, we might wonder why a substantive conception of the good is thought to be necessary to an understanding of law as a moral idea. The liberal ideal is one which gives each person a separate voice in the determination of social good. Liberals have thus tended to regard ideas of virtue as being rooted in the experiences and understandings of individual thinkers who possess the ability to reflect critically on their experience. Even if we can correctly assume that individuals in general possess the intellectual wherewithal to transcend their sectional interests in coming to an understanding of social good, the presence of uncontentious standpoints on matters such as equality and justified entitlement could only emerge as a contingent feature of moral experience. The idea of well-being in Protestant political thought comes to us not as some generalised idea which can be contemplated independently of individual wills and desires, but as something which can only be defined by each person in the light of one's own experiences and needs. It is perhaps the fact that individuals may develop openly distasteful, odd or destructive preferences which prompts the search for 'external' standards, anchored not in the circumstances of individual lives and choices but in some abstract realm of moral value. It is then tempting to regard a person's interests as a matter of what that person *would* recognise as

²² See N. Simmonds, 'Rights at the Cutting Edge', above n.11, at 129.

being good for one if placed in appropriate conditions or blessed with certain insights into the rational or the reasonable.²³ Might we not seek to explain the differences in the way each person values things by the presence of bias and ignorance brought about by purely environmental factors?

As Raymond Geuss has observed, liberal thinking is a form of thinking ‘that needed to be embedded in a more encompassing form of reflection.’²⁴ For liberalism is sustained by a Protestant ethic which views moral experience as fundamentally variegated and uneven. We can then perceive idealism as attempting to flatten out our moral life into a level perspective for the application of moral ideas to all; and as therefore constituting a shift away from a form of reflection in which liberalism can survive as a meaningful ideal. Judgments about what is in a person’s interests become *counterfactual* judgments,²⁵ which lead us to a reformulation of the central question of political theory: we ask not, ‘how can each person’s interests be reconciled with those of others?’, but ‘under what conditions is an individual’s own assessment of one’s situation definitive?’ A person’s entitlements will come to be regarded increasingly as a matter for collective determination and enforcement, rather than an individual’s control over aspects of one’s moral life.

The form of legal reasoning within the common law does not easily lend itself to such idealistic interpretations. Even the categorical divisions of Roman law, seen by the early modern jurists as the apogee of rational legal order, were never more than a heuristic aid for students of the common law despite the various attempts made during the course of the seventeenth and eighteenth centuries to offer Romanist explanations of the sources of law.²⁶ It was perhaps the failure of such efforts to ground common law reasoning in general categorical principles that led Stair in 1681 to proclaim, somewhat disingenuously, that:

“there are not wanting of late of the learnedest lawyers, who have thought it both feasible and fit, that the law should be formed into a rational discipline, and have much regretted it hath not been effectuated, yea scarcely been attempted by any.”²⁷

Such efforts at rationalisation of the law were, however, associated more with widespread dissatisfaction towards the rather cumbersome and irrational collection of rules and procedures which characterised the common law form of pleading than with any sweeping moral or political concern with the form of the legal order as a whole. It was thus with the clarification and simplification of writs and pleadings, rather than the desire to understand writs and pleadings as more mundane instances of some greater principle,

²³ See the excellent discussion in R. Geuss, *History and Illusion in Politics* (Cambridge, Cambridge University Press 2001) ch.2.

²⁴ *ibid.*, 104.

²⁵ *ibid.*, 101.

²⁶ Of these oft-repeated endeavours, perhaps the best known (though not the most successful) was that of Blackstone, principally in the beginning of Book I of his *Commentaries*: for an informative discussion see M. Lobban, *The Common Law and English Jurisprudence 1760-1850* (Oxford, Clarendon Press 1991) ch.3.

²⁷ Stair, *Institutions of the Law of Scotland* [1681], I.i.17.

that common lawyers principally concerned themselves. The notion of a 'right' was thus something of political significance, but little legal importance, in the history of the common law. For the idea of a right to be meaningful, I have argued, it must operate within a context of fairly stable and precise rules through which rights are identified and delimited. Yet the idea of law as a system of identifiable rules is a relatively new one: until well into the nineteenth century, the common law was conceived principally in terms of remedies and procedures in which the courts could hear disputes and offer just solutions to any wrong which had been correctly presented. It was thus the *form* in which the wrong was presented to the court, rather than the establishment of juridical proofs which lay at the centre of legal thought. The law was viewed as a reasoning process channelled along strict lines governed by the system of writs. The reasoning involved was thus not source-based, but rooted in 'reason', understood in broad terms as a wise and reasonable approach to life. Judges could thus draw their deliberations from ordinary experience and common sense, as well as considerations of policy, philosophy and a pragmatic sense of what was convenient or justified in the circumstances. A legal action was defined as the lawful demand of one's right or due; 'yet this right was undefined in terms of rules. Right was only defined in the broadest sense. . .'²⁸

Viewed in terms of the distinction between objective and subjective right (*ius*), such a notion of legal right is extremely difficult to pin down. Despite the obvious overtones of objective *ius* present in the notion of 'one's due', the common law possessed no conception of individual entitlements as defined in terms of general principles preceding the settlement of cases at law. The same ambiguity frustrates any attempt to cast this classical common law understanding of rights in terms of the distinction between the Will and Interest theories. The idea of demanding one's right involved outlining an interest which had been violated, and which merited a remedy in law. Such a view might seem to favour the Interest theory, since the basis of a legal claim did not essentially concern the exercise of individual powers over another's duties, but rather interests which had been harmed. The Will theory depends upon certain formal characteristics of the modern legal order: principally, it depends upon the conception of law as a body of rules and standards which identify an individual's obligations under the law, as well as relatively clear rules establishing areas of autonomy consisting in claims against interference or for assistance, liberties to engage in or refrain from certain actions, immunities against arbitrary changes to one's established entitlements as well as powers of control in relation to those entitlements.²⁹ Lacking anything approaching these ideas, the classical common law could offer no basis for an understanding of rights as protected choices. Yet neither is there any real conformity with the notion of rights as constituting protected interests; for the procedural emphasis of the system of writs and pleadings made it impossible for plaintiffs to define their rights a priori in terms of established rules or principles. Plaintiffs did not enjoy claims to the

²⁸ Lobban, above n.26, 62. I am much indebted to Lobban's book for its rigorous clarification of these sometimes confusing and contradictory aspects of post-Enlightenment common law thought.

²⁹ I refer, of course to Hohfeld's 'fundamental legal conceptions'. I examine each of these conceptions in more detail below.

enforcement of their rights, but rather undertook actions for a remedy in respect of wrongs which had been done to them. The Interest theory of rights can thus be seen to depend upon the same formal conditions as the Will theory.³⁰

This period in the history of the common law is nevertheless of some importance for an understanding of the development of the idea of a right within legal thought. For the emphasis on remedies reveals the common law as a form of moral association underpinned by an idea of the moral life of the polity as an unstable and deeply irregular landscape in which no level horizons for moral reflection are possible.³¹ The virtue of the common law was its *adaptability*: each case to come before the courts was different, because no two cases arose in exactly the same way, or in exactly the same circumstances, and thus each case demanded precise and careful deliberation rather than a tailor-made judgment fashioned from stock rules and principles. The strict rules of the system of writs were thus designed to set each case within specific limits, to render their peculiar features intelligible to judges who could then bring their amassed experience to bear on *those* facts. Legal decisions created precedents only informally: not by laying down formal rules to be followed in the future, but instead by sharpening judicial awareness of dimensions to a problem: 'A clearly focused precedent would articulate the custom [of judges] better than the oral culture of judges, but it was not binding.'³²

The rationality of the common law thus consisted in the thought that the distinctive features of each case demanded close and separate attention, and could be brought within the scope of broad and categorical principles only at the risk of distortion. Yet the legal order was ultimately to undergo a shift away from a concern with procedures and remedies, and towards the notion of a system of established *rules*. This shift is reflected in the legal writings of the time: Thomas Starkie, in his Inner Temple lectures of 1834, argued that 'The first and great business of the law is, to define rights and correlative duties of all kinds, whether they be public or private . . . [I]n our law, the extent of the right is limited and defined by the extent of the wrong.'³³ Legal rights thus gradually came to be identified not with broad and variable arguments concerning the justice of a state of affairs, but with an established body of rules which 'limited and defined' the bounds of permissible and impermissible action.

The continuing association of rights with remedies is of great importance in this context. The traditional common law had viewed rights as a claim for

³⁰ That is, conditions which make the idea of a subjective right intelligible. Common law actions for remedies thus could not be analysed as the application of legal protections for Hohfeldian claim-rights established in advance of their being violated.

³¹ I borrow this image from Nigel Simmonds (above n.11).

³² Lobban, above n.26, 79. See also D. Lieberman, *The Province of Legislation Determined: Legal Theory in Eighteenth Century Britain* (Cambridge, Cambridge University Press 1989). As a number of common law historians and scholars have pointed out, a formal and developed notion of *stare decisis* depends upon an effective system of law reporting wherein the reasoning of judges is recorded and preserved.

³³ *Law Examiner & Law Chronicle* iii, 172-73, quoted in Lobban, above n.26, 187.

the redress of a wrong which lay effectively in private hands. The law provided mechanisms for redress, through the establishment of complex forms of action, and it was up to each litigant to pursue a claim by selecting the relevant form. The shift from a remedies-based conception of law to a rule-based one represents a clear shift from a customary to a statist view of law: the realisation had begun to dawn on eighteenth and early-nineteenth century lawyers that the law had become more than a system for redressing wrongs, but could be seen in more general terms as an instrument of social engineering and rule-based governance.³⁴ The law thus came to be represented as a system of imposed rules and sanctions constituting public and general standards of conduct, and a distinct system of doctrinal ideas and principles through which individuals could organise and structure their private lives. Rights featured importantly within this bifurcated notion of law as private instruments through which individuals could exercise control over aspects of their lives, by making decisions in relation to the doctrinal rules as they affected their interpersonal relationships: by waiving or asserting claims, exercising powers and waiving immunities, and by ordering their affairs within areas of liberty between the rules and obligations. The common law thus operated in a context in which rights were not identified with, but opposed to, collective measures for the protection of individual interests.

Rights, Powers and the Will

The modern lawyer confronts a legal order of increasing size and complexity. Within this complex world, the categorical divisions adopted or presupposed by the practitioners of the nineteenth century seem inevitably contrived and increasingly out of step with reality: for we are used to moving within a legal world in which complicated statutory regimes fuse with developed systems of doctrine to regulate private and commercial contracts, compensation for personal accidents, the use of property and the like, and we accept such amalgamated approaches as an unproblematic aspect of legal life. The idea of the law as embodying two distinctive but interconnecting realms, one system of imposed rules establishing public standards of conduct, and a separate system of evolved doctrinal rules for the regulation of private life, comes across as artificial when measured against today's complex realities; whereas the view of law as a single, coherent and integrated system of principles addressed at once to the large concerns of state and the small detail of private lives, strikes the mind as a plausible and powerful insight.

Fuller was one of many jurisprudential writers who were aware of this general shift in legal thinking, and who sought either to understand it or to contain it.³⁵ Fuller had attempted to separate those cases in which a litigant has some meaningful control over the way in which her interests and claims

³⁴ See Lobban, above n.26, 204, and S.F.C. Milsom, 'The Nature of Blackstone's Achievement' 1 *OJLS* (1981) 1.

³⁵ See for example M Cohen, 'The Basis of Contract' 46 *Harvard L Rev* (1933) 553; R. Pound, 'The End of Law as Developed in Legal Rules and Doctrines' 27 *Harvard L Rev* (1914) 795; more recently Milsom, above n.34, and D. Kennedy, 'From the Will Theory to the Principle of Private Autonomy: Lon Fuller's "Consideration and Form"' 100 *Columbia L Rev* (2000) 94.

are represented, from those in which polycentric features of the case make the litigant's participation in legal argument meaningless. Yet Fuller was also aware that '[t]here are polycentric elements in almost all problems submitted to adjudication . . .' and that the distinction between polycentric questions and 'straightforward' doctrinal problems 'is often a matter of degree'.³⁶ Since adjudication operates in a context wherein doctrinal rules are increasingly underpinned and supplemented by black-letter provisions (and where those black-letter provisions are fleshed out and interpreted against the background of established doctrinal ideas or general conceptions of 'right'³⁷), it becomes difficult to point to cases in which a ruling as to the rights and duties of the parties will have no wider, unforeseen impacts upon the treatment of future cases. Indeed, as Fuller noted, *any* system of laws containing a reasonably robust doctrine of precedent will render the separation of public and private spheres in this way largely artificial.³⁸ Fuller nevertheless believed that the point of such a separation was intelligible: 'It is a question,' he said, 'of knowing when the polycentric elements have become so significant and predominant that the proper limits of adjudication have been reached.'³⁹

I have suggested that the distinction between the Will theory and the Interest theory should be understood as proposing differing conceptions of the boundary of private law. Whereas the Will theory sits comfortably within a theory of legal reasoning which minimises the polycentric dimensions of adjudication, the Interest theory propels polycentric problems into the very heart of legal reasoning by defining the right in terms of the good. Such a move, I argued, is incompatible with the liberal presuppositions of much of our legal and political practice. The modern political world is one in which morality is viewed in a context of plural conceptions of the good rather than of monolithic values. Having given up the idea of a pre-civilised state of nature structured by natural patterns of right and duty, we confront a social world characterised by perpetual and irresolvable tensions. Little can be inferred about such a world from man's 'natural' state beyond a few general truths about the desirability of peaceful society; but there is nothing about those general truths which could serve as the basis for detailed suggestions about the form that society should take. Since it is in relation to the basic goods that add meaning and structure to our lives that endemic disagreement arises, any attempt to define rights to such goods in terms of the interests we have toward them will inevitably fail to resolve widespread tensions by an appeal to rights themselves: 'rights' will have become, in effect, a mere adjunct to a broader theory of the good which exists in abstraction from the beliefs and desires of ordinary people.

Liberalism stands in some tension to forms of reflection which seek the resolution of conflicting moral standpoints in sweeping theories of the good.

³⁶ Fuller, above n.12, 397.

³⁷ I refer, of course, to the Human Rights Act 1998.

³⁸ See also Cohen, above n.35. Any system of private rights depends, of course, on collective recognition and enforcement. It is therefore difficult to imagine a form of legal order in which there would be no public interest in the content of private law rules. The issue for the Will and Interest theories, however, concerns the peculiar form in which such interest is (or should be) manifested.

³⁹ Fuller, above n.12, 398.

For we can view perpetual disagreements concerning the good not as problems to be ironed out by some more encompassing moral theory, but as features which define the human condition. Human societies can then be looked upon as evolved responses to the problem of disagreement, and as thus constituting a source of moral insight into human nature. The Will theory of rights offers a particularly attractive way of looking at the liberal political order. If there are no natural or predetermined social roles and hierarchies, but simply a set of problems up for collective determination, then no aspect of individual lives is in principle off limits to public scrutiny and regulation. It follows that a meaningful doctrine of individual rights must be capable of shielding individuals from the intrusive gaze of the state (as well as intrusions by other people) by establishing and preserving blocks of private life over which those individuals exercise a measure of normative control. Without the presence of legal claims against interference, liberties, immunities and so forth, the notion of individuals having interests distinct from those of the collectivity would have become a comforting fiction.

The notion of a 'right' is thus connected in liberal thought to the idea that individuals possess certain abilities and characteristics which do not belong to society, and over which society cannot make unlimited claims. The powerful intuitions which drive both the Interest theory and the Will theory therefore reflect an association between rights and interests which is indirect rather than central. But whereas the Interest theory downplays this possessive quality of rights by viewing interests as the product of collective determinations, the Will theory more clearly articulates the justificatory basis of rights: so long as individuals retain a sphere of personal autonomy in which they can resist injunctions to act to their detriment in favour of the common good, they remain free to formulate and pursue courses of action whether or not such actions conform to broadly conceived collective interests. It is this sort of freedom which underpins the liberal ideal of the individual as a thinking being with a level of self-control, rather than a mere drone existing for the benefit of the hive. Rights are thus not something which each person has insofar as society grants them out of its aggregative and distributive projects, but something which a person owns.⁴⁰ The recognition of such rights is, of course, underpinned by mechanisms of collective enforcement, so that the boundary between the public and private realms cannot be drawn with any rigidity. The integrity of rights is hence best preserved by the development of a body of clearly defined doctrines and rules which can be contemplated and applied in relative isolation from more general investigations into the political good.

It was Hart who first clearly traced the different ways in which a person's entitlements can serve the right-holder's autonomy.⁴¹ Drawing on Hohfeld's analytical distinctions, Hart observed that the unifying feature of the four Hohfeldian entitlements, the reason we are tempted to refer to such distinct

⁴⁰ See C.B. Macpherson, *The Political Theory of Possessive Individualism: Hobbes to Locke* (Oxford, Clarendon Press 1962).

⁴¹ H.L.A. Hart, 'Definition and Theory in Jurisprudence' in *Essays in Jurisprudence and Philosophy* (Oxford, Clarendon Press 1983) 21-48, 35-36. See also Hart, 'Legal Rights' in *Essays on Bentham* (Oxford, Clarendon Press 1982) 162-93. Hart's suggestions have been taken up and further explored and refined by later writers. See e.g. Simmonds, above n.11, 218-29.

ideas as 'rights', is that '. . . in all four cases, the law specifically recognises the choice of an individual either negatively by not impeding or obstructing it (liberty and immunity) or affirmatively by giving legal effect to it (claim and power).'⁴² Hart's insight serves as an important reminder of the connection between rights and remedies: by focusing on a person's ability to make normative determinations about his situation, or to pursue his projects unimpeded by the choices of others where he lacks the ability to make positive determinations, the Will theory successfully depicts the way in which rights operate within the established and stable boundaries of relatively precise doctrinal rules. The focus of adjudication tends to be upon the historical circumstances of each case, addressing broad and polycentric questions of right and social good comparatively rarely. The common law approach to justice thus reflects a concern with the way in which relations between the parties came about. It is therefore within the context of fairly focused doctrinal and circumstantial concerns, rather than in a spirit of wide-ranging political and social inquiry, that questions of legal right are typically addressed.

But why, a defender of the Interest theory might respond, should we construe those more focused concerns as concerns with the right-holder's *choice*, as opposed to his wider *interests* (say, in having his decisions taken seriously or in being a proper focus of respect in a context where other interests and values are at stake)? In dwelling on the idea of choice, do we not simply presuppose the validity of the Will theory over the Interest theory? After all, we can point to the presence of legal rights – such as employment rights – which, though unwaivable by the right-holder, nevertheless seem to protect important aspects of the right-holder's wellbeing, and are thus ordinarily and plausibly thought of as constituting 'rights'. Statutory rights against unfair dismissal, or racial prejudice in the workplace, or the entitlement to a minimum wage, undeniably enhance the lives and wellbeing of individuals in various important respects, whilst at the same time reducing the scope of those individuals' options and choices over their normative situation. Indeed, it is precisely the inability of the right-holder to bargain away such rights that constitutes their importance as modes of protection. Moreover (it is said), it is not at all clear that the possession of a right invariably depends upon the right-holder's competence to make effective normative choices. The Will theory, it is said, excludes the possibility that children, mentally incapacitated or dead people can hold rights; and yet it is just in the case of inarticulate or vulnerable persons that rights seem to be most important and uniquely apt as a means of protection.

If we are to treat this accusation as more than a humdrum reminder of the fact that our ordinary intuitions about rights can be regimented, conceptually, in different ways, then we must trace the Interest theorist's objection to an evaluative standpoint which attaches significance to the possession of a right even where such possession advances some aspect of the right-holder's wellbeing at the cost of personal autonomy. In that respect, objecting to the Interest theory on the ground that it makes it difficult to preserve the relationship of the common law system of adjudication to its own past, may seem a somewhat blunted response: the common law can be presented as a

⁴² Hart, 'Definition and Theory in Jurisprudence', above n.41, 35.

fully coherent and linear tradition of thought only by resorting to distortive analytical abstractions. Should the implications of the Interest theory lead to the abandonment of deeply ingrained understandings and intuitions about the common law system, the need to preserve traditional understandings of right-based reasoning must therefore find expression in values other than that of tradition itself. I have suggested that a conception of rights which emphasises individual interests at the expense of autonomy should be rejected due to its association with fairly objectionable, if currently fashionable, forms of political idealism. Yet the Interest theorist may wonder whether idealism *is* all that objectionable when compared with the political assumptions of the Will theory. For might we not conclude, on reflection, that it is more important to establish solid protections for aspects of individual wellbeing than to place ultimate control over those protections in the hands of individuals who make their choices in the grip of market forces?

I think there are at least two powerful responses to these objections. These can be stated succinctly, but require considerable elaboration in order to serve as defences of the Will theory:

1. It is central to the liberal idea of politics that judgments about a person's interests are, as far as possible, judgments *of* the person whose wellbeing is at stake. In a large community where the simultaneous and uncoordinated pursuit of projects demands extensive regulation, this 'possibility' generally takes the form of individual power over the enforcement or waiver of duties established in accordance with legal rules, rather than the ability to determine the content of those rules in any direct way. Situations where *both* the content of the rules *and* the enforcement/waiver of duties created in accordance with the rules lies in the hands of empowered officials could be represented as instances of the possession of a right only by casting doubt upon liberal explanations of the basis of the rule of law.

2. In all situations where we tend to speak of 'rights', the holder of the right manifests *de facto* control over the exercise of the right even where officials retain juridical power over the enforcement of duties. The concerted attempt to remove such vestigial control from individuals would require the development of invasive bureaucratic and administrative forms of regulation which, even if practicable, would thoroughly undermine the community's status as a liberal polity. The character of 'unwaivable' rights is thus considerably misrepresented by the Interest theory.

These issues are best explored in the light of a discussion of legal powers. Recent defences of the Interest theory have correctly centred upon the notion of legal powers as identifying the main area of contention between the Will and Interest theories. For, in the absence of Hohfeldian powers to waive or enforce duties correlative to rights, the right-holder would seem to lack the requisite control over the exercise of the right demanded by advocates of the Will theory. We might thus conclude that a legal entitlement qualifies as a right under the Will theory only where one of the basic Hohfeldian advantages is exercised in conjunction with a validly conferred Hohfeldian power. There is a substantial core of truth in these propositions which reflects importantly on the nature of legal rights. Yet care must be taken not to distort that truth in the form of broad assumptions about the matters at issue between rival theories. As Hart observed, rights can defend the right-

holder's choices positively or negatively. Positive recognition is given to a person's autonomy in the form of claim-rights and powers, through which the right-holder is legally enabled to make autonomous choices which directly affect the scope of his entitlements relative to the entitlements of others. Negative recognition is accorded to choice through legal liberties and immunities: liberties define a sphere of freedom from obligation in which the right-holder can make decisions in relation to possible courses of action; whereas immunities defend the right-holder from non-consensual changes (as well as some consensual changes) to her existing entitlements. The negative/positive dichotomy, though illuminating in many respects, has nevertheless served as an obstacle to the understanding of the complex relationship between these various forms of Hohfeldian entitlement and the possession of legal powers.

In a recent and extremely powerful essay, Matthew Kramer defends a version of the Interest theory on the basis of an analysis of legal powers. To hold a legal power, Kramer contends, a person 'must be legally and factually competent to exercise it; moreover his holding of the power must derive directly or indirectly from some legal norms or decisions.'⁴³ If we regard the main issue between the Will and Interest theories as the presence or absence of such powers in conjunction with various forms of legal advantage, then it may seem as if the Will theory is presented with a rather elementary problem: for we can readily bring to mind examples in which the presence of important rights depends upon precisely the sort of trade-off between autonomy and wellbeing which the Will theory refuses to countenance.⁴⁴ For the efficacy of minimum wage schemes, anti-discrimination measures and the like, as well as the possibility of ascribing rights to children and the insane, depends upon dissociating basic forms of legal advantage from powers relating to their enforcement. Broad political visions aside, the threshold for the adequacy of a theory of rights must include the ability to absorb such insights as part of our collective understanding of what rights are.

The problem of ascribing rights to legally (or factually) incompetent right-holders can be met reasonably quickly, so I propose to deal with that first. Kramer gives the following account of the 'prime point of controversy' between the Will and Interest theories:

"According to the Interest Theory, anybody *X* can hold a legal right irrespective of whether he holds any legal power to enforce/waive the duty that is correlative to the right. According to the Will theory, by contrast, *X* holds a legal right

⁴³ M.H. Kramer, 'On the Nature of Legal Rights' 59 *CLJ* (2000) 473-508, at 474. (Hereinafter, *Nature*.) See also Kramer, 'Rights Without Trimmings' in Kramer, Simmonds & Steiner (above n.3).

⁴⁴ I am aware that the following pages might give a misleading impression of the enormous analytical strengths of Kramer's essay. Much of the following argument is inspired or provoked by that essay, and it is indeed the potency and range of Kramer's arguments against the Will theory which demand a fairly substantial engagement with the detail of Kramer's arguments.

if and only if he holds the legal power of enforcing/waiving the correlative duty.”⁴⁵

Such a formal characterisation of the Will theory might seem somewhat surprising to someone in possession of Hart’s distinction between positive and negative protection; for that distinction rested not on any formal characteristics rights may be presumed to possess, but on a substantive vision of the purpose which rights might be thought to serve. Immunities and liberties, for example, do not inevitably depend upon the right-holder’s possession of legal powers of waiver in order to discharge their function of protecting aspects of the right-holder’s autonomy. All the same, we might interpret ‘right’ in this context as specifically designating Hohfeldian claim-rights, in which case the presence of legal powers of waiver may well be a substantive condition for the conferral of rights.⁴⁶ Kramer’s argument then proceeds to outline the attractions of the Interest theory over the Will theory in this context:

“Because infants and mentally incapacitated people and animals and dead people are not competent to demand or waive the enforcement of their entitlements, their holding of genuine legal rights cannot involve their holding of the legal powers to demand/waive the effectuation of those rights. There is no alternative to placing such powers in the hands of others (parents, guardians, public officials and so forth), if infants and mentally incapacitated people and animals and dead people are to have any rights at all. Hence, given that there is no alternative, and given that the conferral of genuine legal rights on the aforementioned people and creatures is highly desirable, the fact that other people exert the powers to enforce/waive those rights is not a cause for keen consternation.”⁴⁷

A short response is this: if such right-holders lack the ability to exercise legal powers, then they *ipso facto* lack the ability to wield and assert legal claims, in the form of Hohfeldian claim-rights. Now, such an implication is not immediately disturbing to the Interest theorist since he will not regard the ability to *wield* legal claims as a precondition for the possession of such claims. Claim-rights are accordingly treated by Interest theorists as mere markers of protected domains, whether of choice or of interest. A legal power must thus be called into play to make legal determinations about how a domain of legal protection is to be defended or altered. A major reason for the discomfort we feel over the ascription of rights to incompetents derives from assumptions about the nature of interests: though claim-rights are (for the Interest theorist) markers of interests, the notion of ‘interest’ is defined in a way which might not coincide, or may coincide only contingently, with the

⁴⁵ Kramer, *Nature* 474.

⁴⁶ In fact, Kramer regards the presence of legal powers as a conceptual, rather than a substantive condition, for the conferral of rights. My reasons for disagreeing with this suggestion will become evident in what follows.

⁴⁷ *ibid.*, 478. See also N. MacCormick, ‘Children’s Rights: A Test-Case for Theories of Right’ in *Legal Right and Social Democracy* (Oxford, Clarendon Press 1982) 154.

professed interests of the putative right-holder.⁴⁸ The possession of a claim-right would not by itself allow the right-holder (incompetent or otherwise) to make any significant legal decisions about how his interests are to be protected, traded or modified (or abandoned) in a given context. The most that such a right-holder could do is call the claim-right into service in a very limited way (by acting in ways consistent with assumptions that the right will be respected, or by informally resisting intrusions by calling attention to the existence of a specific duty). But the lack of a legal power on the part of the alleged right-holder to make any significant changes to his legal position relative to others, by enforcing or waiving the right against specific persons, is the key reason why we ordinarily feel uncomfortable about ascribing rights in such contexts.

For a claim-right to be operative – that is, to have actual legal effects – it must therefore be associated with a legal power, regardless of who holds that power. A claim-right-plus-power is therefore minimally sufficient for the existence of a genuinely operative legal entitlement against unwanted intrusion (or for assistance). Given that claim-rights can be wielded only through the use of such powers, the situation of the incompetent right-holder is thus misleadingly characterised by the Interest theorist as presenting a rigid contrast with the Will theory. For the question is not one of rigid analytical boundaries (between entitlement and non-entitlement), but of whether the presence of choice in relation to the way in which the right is exercised is sufficiently analogous to the situation of a competent right-holder being in full control of that choice to make an attribution of right plausible in the case of the legally incompetent. A competent right-holder, *X*, rarely wields and asserts a claim-right himself; that business is usually carried out by a lawyer acting as *X*'s agent in accordance with *X*'s instructions and consent. A situation in which an incompetent right-holder, *Y*, is consulted about changes to his legal position brought about by the use of a legal power held by *Z* (a parent or guardian, say), and which is used by *Z* in a manner consistent with *Y*'s expressed wishes, is thus sufficiently analogous to the position of the competent person as to warrant an attribution of right to *Y*, albeit that the right is in fact exercised by *Z*. We might, of course, question on more general grounds whether claim-rights can be exercised on behalf of another and yet still be referred to as that person's claims; but since the Interest theory's ascription of rights to the incapable depends upon so doing, the Interest theorist ought to hope that we can.⁴⁹

In contrast to cases which closely resemble *X*'s position on the matter of choice, we can imagine instances in which *Z* exercises *Y*'s claim-rights with no regard for *Y*'s own stated preferences (and perhaps also without regard for *Y*'s interests at all). In such cases, it is clear that the analogy with the

⁴⁸ Indeed, depending on further assumptions about collectively determined interests, the interests protected by the right may even differ from the right-holder's *actual* interests (irrespective of the latter's own views).

⁴⁹ At least in the case of the very young, it is not clear that we can. As Hart pointed out, moral conditions placed upon our actions towards children do not seem to be based upon respect for rational autonomy, but on a generalised concern for aspects of their welfare. This is not a *reason* for the assertion that children do not possess rights, but an *explanation* of a proposition which (if it holds at all) holds independently of the Will theory.

possession of a right by *X* has become broken-backed: even if *Z* overrides *Y*'s known preferences with good reason or good intention, or if *Y* is incapable of expressing preferences at all, the lack of control by *Y* over his legal situation is so pronounced as to undermine the rationale for characterising *Y*'s position as one involving right rather than regulation on the basis of a concern for welfare. In relation to the legally incompetent, therefore, we are presented not with a rigid categorical division between cases of right-holding and cases of non-right, but rather a continuum of possibilities which are more or less like the standard case of the competent right-holder. The Will theory is thus able to explain our commonsense judgments in such situations (and the level of discomfort we feel about the attribution or denial of right) in a more nuanced way than can the Interest theorist.

A quite different, and more complex, problem is presented by rights which seem to trade autonomy for wellbeing in line with broad social policies. A society will engage in such trade-offs where there is a strong reason to believe that individual or collective welfare would be threatened if certain choices were left to the uncoordinated wills of individuals. The withholding of benefits or the application of penalties on grounds of race is one of a number of issues recognised as being too important to both individual wellbeing and to the community's collective self-image to be left to private decision. Where a person chooses to sacrifice such interests in order to gain some other benefit, or to avoid some threatened consequence, he not only acts against his own welfare, but also erodes and undermines his community's values. There is nevertheless a good deal of sense in the notion that such protections should be counted among a person's *rights*. Minimum wage laws, for example, undoubtedly confer complex entitlements on workers (including claim-rights to be remunerated above a certain level), but simultaneously restrict the aggregate of choices open to a person to act within the law:⁵⁰ certain otherwise valid contracts of employment can no longer count as legally enforceable bargains.

Whilst it seems intuitively correct to speak of rules of this kind as conferring rights, such regulatory regimes vary the extent of individual entitlements in many ways, both positive and negative. By altering our perspective, we can examine those effects in different ways: a focus on each person's claim-rights to levels of remuneration from within a given range are naturally presented as affording positive recognition to choices which each person might make to accept offers and make legal bargains within that range. Viewed from the perspective of the consequent reduction in the number of opportunities for meaningful employment, however, we are hardly likely to regard the lack of legal recognition accorded to a person's choice to accept an employment offer in violation of such protective and regulatory norms as a 'right'. We will rather look upon those norms as having conferred extensive disabilities and liabilities along with liberties and claim-rights. The issue of interests is tangential to such questions: our willingness to ascribe rights to people in the light of wide-reaching normative regimes is

⁵⁰ In restricting the scope of legal action, such laws effectively reduce a person's freedom overall: see M. Kramer, *The Quality of Freedom* (Oxford, Oxford University Press 2003).

quite separate from the judgments we make about the way in which those regimes distribute interests and costs.⁵¹

Where rights are conceived as protected choices, it is possible to preserve the intuition that regulatory regimes should be presented as a balancing of rights against wider concerns with welfare, rather than a straightforward balancing of rights against other rights. A concern with individual wellbeing does not invariably take the form of a concern with the recognition of rights. The project of effecting some overall shift in the distribution of benefits across society (such as the institution of a basic level of remuneration) is not *the same* project as balancing the rights of every person to secure and pursue a worthwhile life: it constitutes instead the desire to modify the extent of each person's legal entitlements in order to secure some other social advantage (aside from the protection of rights), such as the workability of a system of protections which would collapse if left to the play of market forces. The Interest theory erodes the integrity of rights by failing to recognise such developments as concerning the boundary between individual rights and collective welfare rather than dealing exclusively with the boundaries between competing rights.

Rights conferred by statute as part of some wider strategy or social goal are rarely considerable in isolation from the complex regulatory mechanisms which guarantee the effectiveness of those rights as demanded by the strategy. It is perhaps the tendency of the Interest theory to look upon such regimes holistically which explains the theory's treatment of regulatory provisions as aspects of questions of right. Regulatory provisions are then capable of being seen as directly defining the boundary between competing rights, rather than as higher-level norms defining the operative bounds within which rights compete. The Will theory, it seems to me, offers a credible and attractive explanation of why the benefits conferred by legislation should be looked upon as giving rise to rights, whilst preserving the distinction between rights and regulation. We have already seen how the trade-offs between autonomy and wellbeing which are typical of statutory regimes of regulation can be presented as refutations of the choice-orientated view of rights only by assuming that the questions raised by regulatory measures must be dealt with on the same level as questions of individual right. It is the failure to see that there are in fact *two* distinct sets of questions involved (questions about how individuals' lives are to be regulated in their own and the general interest, and questions about the effects such regulations will have on the complex web of entitlements which obtain between those individuals) which leads the Interest theorist to suppose that questions of choice are to be disposed of *within* the language of rights (by drawing a distinction between waivable and unwaivable rights) rather than a context in which rights compete with other values and interests. Yet the interplay of interest and choice is perhaps more complicated than we might think.

Let us consider the example of statutory employment rights which protect workers from unfair dismissal on grounds of race, religion, disability etc. If

⁵¹ That is, we can regard a regime *R* as distributing costs and interests in a way which is broadly beneficial (perhaps deeply beneficial) to individuals in society, but as bringing about those benefits at the expense of certain rights. For a sharply different view see Kramer, above n.43, 94-96.

such rights are to operate successfully, it is clear that employees must possess legal disabilities preventing them from alienating their legal protections as part of the normal bargaining process between employer and employee. We can therefore view such protections as helping to define the *kind* of agreement which can exist between the parties to a contract of employment: such agreements will include duties owed by the employer towards the employee which the latter cannot give up or trade for some other proposed benefit. As many defenders of the Will theory have pointed out, however, the inalienability of the rights to which such duties are attached is connected with the issue of waiver only through quite complex chains of reasoning: for although the right-holder lacks the legal power to bring such duties to an end, he can clearly control the application of the duty by deciding whether to sue or not to sue.⁵²

The choice of the right-holder thus remains central in determining when the right will be brought into play in the service of his interests.⁵³ The inalienability of rights addresses a separate question from that of identifying the right-holder's interests: the permanency of such rights secures anterior goals, such as the workability of a system of protections, which define a context within which an individual can make choices about his own normative situation. By equalising the bargaining powers of the parties, a system of inalienable rights (though clearly reducing the number of choices open to each person) creates conditions in which an individual's remaining choices carry meaningful weight. Rights of this kind are therefore not vastly different in kind from ordinary contractual rights, but only in degree: statutory regimes ensure the establishment of rights which, except for inequality of bargaining power, might have come about through contractual negotiation in the usual way. Such rights will tend not to affect the scope of a right-holder's choices significantly (unless the right-holder adopts strange preferences such as the preference to be discriminated against on racial grounds), but will merely render certain operations with the right, such as bartering, unenforceable. Yet even in the case of ordinary contractual rights, rules such as those requiring valid or additional consideration operate to

⁵² See *e.g.* Lucy, above n.4, 195. The right-holder may, of course, possess some residual powers to bring the duties to an end: for example by serving notice on the contract as a whole, or perhaps by electing not to raise an action upon the duty within the scope of prescribed time-limits. We might think of the alienating/not-exercising distinction as pointing to different ways in which a right can be waived.

⁵³ There will be occasions, however, on which this is not the case. Under Art.44 ACHR, for example, any person can institute proceedings against a perpetrator of torture, irrespective of the victim's consent. Examples of this kind will be comparatively rare, but in any case raise a separate issue which does not threaten the Will theory: here, a certain person *X* has a claim-right, but the legal power associated with the right is vested generally in a group of persons (including *X*). Accordingly, *X* possesses no immunity from non-consensual changes to his legal position due to the exercise of that power by others. The position is thus characterised not by *X*'s possession of a claim-right-and-disability, but by his possession of a claim-right-and-power-and-liability. Will theorists (and many Interest theorists) will be uncomfortable with the presence of such a liability, but that discomfort does not undermine the validity of the Will theory's analysis of that situation as one in which *X* has substantial though not total control over a claim-right.

place restrictions upon the circumstances in which rights can be alienated, waived or enforced. The inability to bargain away the right at the outset – as opposed to waiving performance of the duty when breached – will represent no significant loss of autonomy to the vast majority of right-holders, and it is thus appropriate to refer to such schemes in a general way as being concerned with the creation of rights.

While such explanations have a good deal of force, we might nevertheless harbour some suspicions that the notion of 'waiver' cannot be so easily restricted to a decision over whether to sue: the distinction between alienability and enforcement might prove important in some respects, but (we might feel) it is hardly absolute. Where officials of the state retain the power to waive or compel performance of duties on their own initiative, or where statutory duties can be modified or eliminated only through further acts of amending legislation, the gap between alienability and enforcement becomes somewhat tenuous and unreal. Such apparent problems for the Will theory evaporate, however, if we view the boundary between right and non-right as a matter of evaluative judgment rather than analytical stipulation. Rights can be associated with powers of waiver in numerous possible ways, each with its distinctive effects upon the right-holder's choices. In some of these cases, the right-holder's vestigial control over some legal advantage will incline us to regard that advantage as a 'right'; in other cases we might view the holder as having insufficient power to decide the application of advantageous legal rules to his own situation, and thus regard his position as regulated rather than entitled. There is no reason to suppose that such judgments are governed by strict semantic criteria, or by values which apply evenly to every case.

Where Interest theorists speak of 'unwaivable' rights, they presumably have in mind situations where the right-holder enjoys the benefit of some advantage (a claim-right or immunity, say) unadorned by any power to determine the way in which that advantage applies to her situation. In such cases, we would then look upon the person concerned as having a right, coupled with a *disability* to deflect the benefit of the right. Might we not then view anti-discrimination rights, such as those which function in the employment context, as Hohfeldian immunities associated with disabilities against the deflection of benefit?

I think not. There are certainly conceivable situations in which the subject of some legal advantage retains no vestigial power whatsoever to deflect the unwanted benefits of that advantage: we might, for example, consider the position of a free-thinking member of a caste-system, who may have extensive claim-rights against all physical contact by members judged to be 'inferior' within the social hierarchy, but who objects to the system of rules and regulations which perpetuate such class distinctions. True, such a person could ignore or even encourage infractions of his own claim-rights, but he would be incapable of waiving the duties of the underclass towards him, whom the state may continue to punish mercilessly. Could we really speak of measures which establish categories of unlawful interference in this way as conferring *rights*?⁵⁴ (Suppose the man wished to marry a woman from

⁵⁴ We are particularly likely to answer in the negative if the person encouraged the infractions as part of his efforts to instigate a change in the regime – by

within the underclass, but was prevented from doing so by the presence of such ‘rights’?) In the vast majority of non-oppressive situations, the subjects of legal advantages do not possess such extensive disabilities, but continue to exercise vestigial control over the application of their entitlements as they see fit. The right-holder, in most cases, can *decisively* shape the legal response to her situation by electing to raise an action, or in some cases merely by reporting infractions. Only in cases where the recipient of some advantage has no possibility of shaping the outcome (where, for example, the failure to receive some benefit, or to report the infraction of a duty, is *itself* an actionable wrong) is that recipient wholly devoid of legal powers to determine the application of entitlements to her life.

The vast majority of regulatory systems will therefore invest the right-holder with interstitial powers to enforce or waive correlative duties. It was HLA Hart who first drew attention to the propensity for claim-rights, powers and immunities to establish a perimeter of indirect protection for legal liberties whose exercise might otherwise be rendered worthless: since liberties consist merely in the absence of duties to refrain from performing the permitted action, liberty-holders are quite unprotected from interference by others, in pursuit of their own liberties, which can frustrate and render nugatory such attempted exercises of liberty. The liberty to carry on a business in a hostile marketplace, for example, enjoys considerable protection from established claim-rights against libel, theft, restrictive practices, insider dealing and the like. No corresponding attention has however been paid to the way in which a person’s established rights can protect and define interstices of *power*. For each person typically possesses an array of immunities against prosecution for failing to raise or pursue actions established as possibilities by statute, coupled with general immunities against alteration of one’s existing legal entitlements by the action of officials seeking to raise or abandon legal actions on one’s behalf. The law also confers myriad bilateral liberties which make possible appropriate choices about how each person will manipulate their other entitlements in the world of social interaction, by asserting and pressing claims, threatening legal action, or securing alternative benefits by agreeing to waive an action at a particular point in time.⁵⁵ The

encouraging inter-class contact, say, and in initiating such contact as a means of garnering signatures on a petition or for the purposes of organising a pressure group.

⁵⁵ It will be clear from my analysis of vestigial powers that I disagree with Nigel Simmonds’s fascinating suggestion that Hohfeldian claim-rights form a specialised kind of power within Hohfeld’s analytical scheme, distinct from the generalised notion of power which we know by that more familiar name. This specialised form of power is exactly a power over the enforcement or waiver of duties. (See Simmonds, above n.11, 224). It is conceivable that a system of legal regulation might confer claim-rights (say) on individuals unadorned by any Hohfeldian powers of waiver at all (whether possessed by that individual or by anyone else). In such cases, duties would have been established to which no legal remedy corresponds. Thus, to the extent that it is conceivable that duties might exist to which no enforceable penalty is attached, naked claim-rights of this kind emerge as possibilities. But then in what sense would these amount to genuine *juridical* relationships, as opposed to aspirational expressions of policy or best practice? It is perhaps similar thoughts which lead Hart to remark that ‘If there are legal rights which cannot be waived, these would need special treatment.’ (Hart,

state's regulatory mandates will thus almost inevitably confer significant and important interstices of power.

The question of the exercise of powers in relation to individual entitlements is a complex and important one. The Will theory reflects that importance and that complexity very well; for it is surely in connection with the exercise of control over aspects of one's normative standing that the notion of a right is most naturally and appropriately used. The foregoing analysis suggests just how subtle and variegated such control can be. The Will theory of rights does not dogmatically presuppose the existence of a 'right' whenever immunities, claim-rights or liberties are brought within the scope of a legal power; for the notion of a right, like the question of control, is an evaluative one which can change within a context of varying degrees and kinds of vestigial power. The notion of a 'right' is not one which derives its shape from the drawing of convenient analytical boundaries, but from judgments whose conceptual associations are inseparable from the *point* of reflecting on the standing of individuals vis-à-vis their fellow men and the state. If the point of making such judgments reflects a concern with delineating those areas in which a person's life is free of collective control, it becomes clear that not every conferral of a legal advantage will amount to a 'right'. Certain immunities, for example, will qualify as rights even where the holder retains very little vestigial power over their application, since they provide important forms of negative protection to personal autonomy. (We can view immunities against wrongful arrest and prosecution as being of this kind.) Other readily imaginable unwaivable immunities, on the other hand, do not by ordinary standards count among a person's rights. The rules of a respected scholarship programme, for instance, may provide that applications from scholars over the age of 30 will not be considered. If I am above that age, I am legally immune from having my existing entitlements enriched by the exercise of any power of the awarding body in deciding to whom the scholarship should go. The relevant difference between these two immunities lies in the point for which they were conferred: my immunity from arbitrary arrest exists to protect me from intrusive, autonomy-challenging powers of state officials, whereas the point of the immunity conferred by the scholarship rule is not to protect me, but to establish parameters within which others can compete on a level playing-field. Whereas the first is focused on *my* choices, the latter is the consequence of the choices and actions of others.⁵⁶

above n.41, 36. For a quite different and detailed treatment see Kramer, *Nature*, 481-95.)

⁵⁶ The same holds true of the other Hohfeldian relations. A power to waive or enforce contractual rights is naturally spoken of as a right since it allows me to make significant determinations about my legal relationship with others. But if I crash my car into another's, I also exercise a legal power since I thereby alter my legal relationship vis-à-vis that person (and anyone who is injured as a result of the crash, and presumably his and my insurance companies). It would be ludicrous to speak here of a *right* to crash my car, since crashes are generally *accidents* rather than the outcome of a deliberate choice: the power here exists as a mere consequence of unintended action, not as a legal protection for an authoritatively recognised choice.

The Empire of Interests

Modern legal and political thinking is dominated by the idea that rights protect aspects of individual wellbeing. The value of autonomous decision represents just one interest to be weighed and balanced against others on the dominant view. The conception of interests at work in such thinking is thus an idealistic one in which a person's own prioritising of his interests is 'hardly dispositive' in relation to the legal enforcement or relaxation of interest-protecting duties:⁵⁷

"If all or most of the members of a certain class of people lack the mature or informed judgment that is needed for an accurate appraisal of their own fortunes, then the fact that most of them view a particular practice as undesirable is not determinative of the way in which that practice should be construed. To whatever extent is appropriate, the Interest Theorist's conception of well-being or interests is objective rather than accommodatingly subjective. Although someone cleaving to that conception will typically defer to the evaluative judgments that are reached by most people in most contexts, he will depart therefrom if he has good grounds for thinking that those judgments are misguided or myopic."⁵⁸

I have argued that such a view of interests entails the rejection of a form of positivism which is essential to any attempt to deal properly with the 'Protestant' political assumptions which characterise the modern age.⁵⁹ Much of the attraction of the Interest theory can be traced to the changing social conditions of the eighteenth century: the rise of capitalism from the ashes of the feudal order, on the one hand, and the enjoyment of considerable periods of relative political stability and moderate government, on the other, created conditions in which individuals no longer viewed the impending loss of control over important areas of life (such as religious observance or political allegiance) as the most pressing social issue. With liberty no longer under sustained threat, political focus shifted from grand political concerns to a concern for certain sectional interests. The focus on particular interests rather than the general political conditions in which those interests are pursued both emphasised liberal disagreement about the good life, and suggested a solution to the problem of disagreement: for we could construe liberal equality as demanding not the freedom of individuals to pursue widely divergent conceptions of the good within a framework of 'neutral' rules, but the participation of each person in the delineation of a shared

⁵⁷ Kramer, *Nature*, 497.

⁵⁸ *ibid.* See also 508.

⁵⁹ One might try to bring such a standpoint in relation to interests within the scope of positivism in other ways: for instance, by recognising that a legal system's Rule of Recognition can absorb such idealistic impulses as *contingent* features of legal practice. This gambit does indeed achieve a resolution of that analytical tension; but it is not clear what such a diluted form of legal positivism contributes to the attempt to understand the distinctive forces of Protestant political theory: positivism, in these circumstances, would have become detached from its historical immersion in robustly liberal values and ideals. (For a broadly similar accusation, see D. Dyzenhaus, 'The Genealogy of Legal Positivism' 24 *OJLS* (2004) 39-67.)

conception of the good distinct from the 'myopic' desires of individuals.⁶⁰ Rights then no longer serve to demarcate spheres of individual activity which are free in important respects from social control, but instead establish focal points for collective concern. Yet a society in which all spheres of personal autonomy have been eliminated or have become the subject of social control, is surely correctly viewed as one wholly devoid of rights.

Kramer has offered an argument purporting to refute this objection.⁶¹ For, he argues, it is entirely possible to characterise the dismal social conditions of such a society without resorting to the proposition that citizens of that society lack rights. Such a society would be characterised in Hohfeldian terms as one in which individual citizens possessed claim-rights but no powers of enforcement or waiver over the corresponding duties: such powers would instead vest in officials of the state to be exercised in relation to those individuals as and when deemed appropriate. This, says Kramer, does not represent a situation in which citizens lack entitlements (for their interests might be served very well by the array of claim-rights they possess), but one in which those citizens lack the ability to make autonomous choices in relation to the extent of their entitlements or aspects of their general situation.⁶² In such circumstances, Kramer argues, the Interest theory is able perfectly well to capture the awfulness of each citizen's plight:

"In these circumstances, [the citizen] would be roughly akin to people who have undergone the complete amputation of their legs (by public officials) but who hold legal rights against interference with their running in cross-country races."⁶³

The point, however, is not whether the Interest theory can aptly convey the awfulness of the plight of citizens within such a society, but what the value of ascribing *rights* to citizens in this situation would be. That rights *can* be so ascribed on the basis of the Interest theory's definition of rights is not in doubt; what is in doubt is exactly what is gained by clinging to that definition in the outlined circumstances. The situation outlined is one in which the lack of choice possessed by each citizen to alter (or preserve) some aspect of her legal situation is the most pressing feature stifling her life. That lack of choice is thus properly focused upon by the Will theory as defining the difference between a person's possessing entitlements and a person figuring

⁶⁰ Nigel Simmonds presents the Interest theory as a (positivistic) response to the failures of Kantian approaches to justice. Despite the reality of these developments, it is doubtful whether the collapse of Kantianism had anything but a very small influence on the development of legal doctrinal thought in the eighteenth century, when compared with the rise of commerce as the principal form of economic activity, and the general tendency towards reform of the system of writs, charted above. These developments were quite independent of the positivism which was emerging as a distinctive force in legal reasoning. See Simmonds, above n.11 *passim*.

⁶¹ In fact Kramer's argument relates only to a society in which all *claim-rights* are controlled by officials of the state. It is unclear whether the same argument would apply to a society in which immunities, liberties and powers had also become absorbed into state control. I will therefore similarly restrict my remarks to claim-rights in the following paragraphs.

⁶² Kramer, *Nature*, 480.

⁶³ *ibid.*

in a community's political thought as a mere pawn in the state's aggregative and distributive projects. To insist that a conception of entitlement is compatible with *both* situations is to rob the notion of entitlement of any conceivable point.

A situation in which citizens' entitlements are controlled exclusively by public officials is one in which each citizen's entitlements are exercised or waived on the basis of collective goals, or on the basis of aggregated beliefs about individual welfare. Whether or not officials openly exercise their powers over individual entitlements in line with proclaimed or established goals and policies (such as the aggressive pursuit of tort-feasors or a hard-lines approach to personal accidents), or whether they genuinely try to exercise those powers in the interest of the individuals concerned, the reality will be that powers of enforcement and waiver will be exercised according to established rules and guidelines which regulate their application. Where officials are charged with controlling the entitlements of large numbers of people, the growth of a body of general rules delineating the exercise of legal powers is inevitable. Particularly in societies which have evolved a doctrine of *stare decisis*, the growth of a body of jurisprudence around attempts to apply powers on a case-by-case basis will seem the only means (aside from explicitly posited rules) by which officials can approach their task in a comprehensible and even-handed way. The growth of a body of jurisprudential principles will represent the expression of shared ideas (perhaps only among the public officials) about how each individual's interests should be ordered and protected. Though each case will throw up distinctive questions, departures from received or established ideas will require separate justification which will itself contribute to the jurisprudential understanding of how individual rights are to be enforced.

Where the enforcement or waiver of a person's legal advantages depends upon general public rules, there will be no obvious way of drawing the boundary between public and private goals. For each person's interests will be given legal effect just to the extent that they contribute to broad public ideas about how an individual's relationships should be conducted. Suppose John contracts with Mary for the supply of 600 industrial widgets to be delivered by a specified date. Where powers of enforcement and waiver lie exclusively in the hands of public officials, what sense would attach to the claim that John or Mary have entitlements under the contract? If Mary fails to deliver by the due date, causing John to incur considerable loss, John's ability to recover would lie entirely within the purview of officials who have the power to decide whether or not to institute legal proceedings. If officials openly decide issues of enforcement or waiver in the light of broad social aims, John's role in litigation is likely to be confined to drawing purported breaches of right to official attention: for to ascribe any role to the plaintiff in the presentation of claims and arguments before the court would be to invest in individuals a considerable level of residual control over the terms in which their claims are presented.⁶⁴ Having been granted official leave to make a

⁶⁴ Such control might equally, if less directly, be manifested in the instructions given by the plaintiff to his legal representative, who both advises and acts on the basis of his client's instructions. See above for further consideration of this point in relation to legally incompetent right-holders.

claim, those individuals would then exercise effective powers of waiver and enforcement through legal argumentation, by emphasising particular aspects of a claim and by downplaying or conceding others.

A state would have powerful reasons for placing such controls in private hands; but the result would be a straightforward collapse of the empire of interests into the republic of choice.⁶⁵ public powers of 'waiver' would amount to little more than a body of doctrinal rules establishing the parameters of permitted action and outlining clear principles of breach of contract, duty of care and the like. The need for official decision regarding the desirability of legal proceedings can then be seen as constituting a form of pre-trial hearing, for the purpose of deciding whether there is genuinely a case to answer in law. A formal requirement of this kind would mark a curious procedural feature of the common law, but would not have much effect on the internal logic of legal pleading: each citizen would still exercise effective control over her range of legal entitlements within the scope of legal rules and regulations. John would have to convince the relevant officials that his claim falls within certain principles and definitions; but thereafter his position as a litigant would not significantly differ from that of a plaintiff at common law. The converse of this situation – where citizens lack any such control – is therefore marked by official control of every stage of legal pleading. Such a situation could only conceivably represent the desire to subject all private enterprise to extensive regulation in the light of collective assumptions about the common good. It is not obvious what evaluative point is served by retaining a reference to 'rights' in these circumstances except to draw attention to the fact that each person's interests are heavily regulated as part of some amalgamated political ideal.⁶⁶

Conclusion

The circumstances of modern political and legal thought in the Western world have propelled the notion of a right into the centre of our social concerns and understandings. This is not an altogether happy situation, but one which is inevitable as long as we continue to view the social world

⁶⁵ See L. Friedman, *The Republic of Choice: Law, Authority and Culture* (Cambridge Mass., Harvard University Press 1990).

⁶⁶ Suppose someone objected to this conclusion on the ground that officials could in fact seek to waive or enforce duties exclusively on the basis of the interests of the individuals concerned, in the light of the facts of each case. Legal advantages possessed by those individuals would then (it might be claimed) enjoy adequate protection whilst acting as markers of private interests distinct from wider collective projects. Two responses are in order: first, it is hard to see how the exercise of official powers *is* compatible with any meaningful distinction between public and private interests. If official decisions about enforcement/waiver are indeed taken outside the scope of fixed rules (which is doubtful), then legal proceedings undertaken by those officials would simply resemble attempts at arbitration within an adjudicative framework rather than the genuine enforcement of legal rights. Secondly, forms of social order might emerge as abstract possibilities without constituting plausible political realities. Aside from the fact that any such regime would be administratively inconceivable, the point of transferring the responsibility for presenting claims and arguments of fact to public officials in this way escapes rational comprehension if the desire really is to protect individual interests quite apart from collective ideals.

through 'Protestant' eyes. Given the irresolvable tensions of a shared social life, it is in the notion of protected choices that we best reflect and give expression to the liberal ideals which underpin the Protestant political order: for the presence of private domains of autonomous decision is both the defining feature of the modern political order and the source of its most pressing problems. The Will and Interest theories of rights can be represented as opposing solutions to the problem of clashing interests. The Interest theory, I have suggested, should be read as the attempt to find some shared moral perspective against which each person's interests can be balanced out or reconciled; but the pervasive disagreements which attach to the articulation of individual interests forces the Interest theorist to seek moral agreement at the level of highly abstract and generalised political visions. Accordingly, the Interest theory is closely associated with forms of moral idealism which ultimately undermine a commitment to the liberal values which define the modern political landscape.

The Will theory, by contrast, attempts to locate private disagreements about the political good within a context of established rules and principles which simultaneously empower individuals to make meaningful decisions about their relationships with other people, and regulate those decisions. The reality of a private sphere within which individuals are free to formulate and pursue their own aims depends upon a body of rules and standards which are relatively immune from revision in the light of sweeping political interpretations. It is thus in preserving and clarifying the boundaries between entitlement and regulation that the significance of rights for the liberal polity is most powerfully explained.

THE FOREST-PATH OF LAW AND THE WOLF-BELLY OF JUSTICE: LEGAL THEORY AND FAIRY TALE

David Gurnham, Lecturer in Law, University of Manchester

Introduction

European Fairy tales have evolved from oral folk tales and represent all manner of social, historical, political and moral purposes and agendas to their many readers and commentators of all ages. Of the many possible reasons for returning to these old stories, mine is that they have something very interesting to contribute to debates on the idea of 'justice'. Contemporary readings do not agree on the moral significance of the genre, although *prima facie* a morally conservative idea of justice is rarely more apparent than in the fairy tales collected by and published initially in early nineteenth century Germany by the Brothers Grimm.¹ This article first rehearses the conservative view of fairy tales, which holds that the difference between heroes and villains are represented by a conceptual distinction between interior (home) and exterior (outside). On this view, villains such as wicked stepmothers and wolves appear as physically and morally peripheral to the societies which they come to threaten. Such a distinction, we may think, prescribes a fairly clear definition of justice and injustice, and at least for Grimm and others who wrote the tales down, this is deliberately so. But of course, the tales were not the original creations of the Grimms, Charles Perrault² or Giambattista Basile,³ all of whom are considered here. The significance of their work in collecting and publishing existing popular peasant folk tales was to make the stories available in a standard written form that was suitable for a contemporary audience. This article argues that reading the tales in the context of this oral folk history reveals a much more complex and uncertain moral landscape. The tales speak of a darker and more interesting view of justice lurking within the apparently secure moral distinction between 'home' and 'outside'. They warn that danger and violence is not conceptually separate to places (and concepts) that we thought were safe. Specifically, the tales narrate the inter-generational conflicts involved in family life and warn about the dangers of child abuse and incest within the home.

The Moral Conservatism of Fairy Tales

Bourgeois values and gender stereotypes

At its most basic, the fairy tale idea of justice apparent from the collections of the Grimms, Perrault and the Italian Giambattista Basile is that in the end

¹ See J.L.C. & W.C. Grimm, *Grimm's Fairy Tales*, trans. L. Crane (Wordsworth Classics, 1993)

² C. Perrault, *Tales from Perrault* (Oxford University Press, 1988); first published in France in 1697.

³ G. Basile, *The Pentameron*, trans. Sir R. Burton (Spring Books, 1957)

everyone gets what they deserve. Reading the tales as an illustration of this principle, it is difficult not to discern a natural order associated with ideas of gender and justice. It is well known that the Grimm brothers collected folk tales for their collection under conditions of economic hardship and French military occupation of their home town of Kassel. Having themselves been born into a frugal Protestant family, the Grimms worked with the belief that the old folk tales represented an ideal German culture – a vital form which they could use to educate the children of the middle classes in the Calvinist virtues of diligence, industry, honesty, order and cleanliness.⁴ Of course, there was nothing distinctly German about the stories – various versions (some of which are examined here) existed all over Europe – but the Grimms believed that the spirit of the German people was captured in “natural and pure forms of culture”.⁵ They feared that this spirit was in danger of being lost underneath the development of the more “refined” culture of the Renaissance and Enlightenment,⁶ and their project of collecting and reinvigorating old folk tales would preserve the purity of traditional German culture.⁷ Zipes’s own interpretation of Grimm is grounded in both history and psychoanalysis. According to Zipes, Grimm’s resurrection of old folk tales forms was “based on a desire to resurrect the authority of their father and his heritage to regain a lost, untarnished home or realm”.⁸ Zipes suggests that it was the loss of the Fatherland following Prussia’s defeat by Napoleon and dissolution of the Holy Roman Empire in 1806 that motivated the Grimms to reflect on the domestic concerns in their versions of the tales they collected. Wounded by foreign occupation of the Empire, the Grimms reflected on the need for a home, loyalty between family relations, the punishment of evil, achievement of contentment through simple hard work, perseverance and an acceptance of the rule of the Father.⁹ In teaching the difference between right and wrong, the tales present an idea of justice as bound up in bourgeois ideals of industry, home and specific gender roles. Thus, Grimm’s tales are patriarchy personified: Father, Fatherland, patriarchal family values. Marina Warner, to whom we shall return, also favours an historical approach to reading fairy and folk tales. She reads the narratives of wicked stepmothers, not as representative of Oedipal anxieties, but as the realistic voice of women of seventeenth and eighteenth century Europe trying to articulate their vulnerability in their specific social and legal system.¹⁰

Zipes regards Grimm’s tales as deeply conservative and as such not entirely suitable for a modern audience.¹¹ Other commentators share the view that gender and justice are inextricably linked in fairy tales generally. Steven Swann Jones argues that the conservatism of fairy tales is demonstrated

⁴ Zipes, J. *The Brothers Grimm: from enchanted forests to the modern world* (Routledge, 1988) p.34

⁵ *ibid.*, at p.32. However, it should be noted that Germany did not exist as a unified nation state until 1871, following defeat of France in the Franco-Prussian war.

⁶ *ibid.*

⁷ *ibid.*, at p.33

⁸ *ibid.*, at p.34

⁹ *ibid.*, at pp.37-9

¹⁰ M. Warner, M. *From the Beast to the Blonde: On Fairytales* (Chatto & Windus, 1994) Chap.14, pp.218-240

¹¹ Zipes, above, n.4 at p.42

through their themes of social instruction and cosmology. The tales are intended to encourage bourgeois values of “industry and moral virtue”, and characters that display these qualities are rewarded with material goods.¹² At the same time, the tales naturalise earthly values by providing an account of how the universe is. Fairy tales often involve the protagonist being called upon to solve some problem, pass a test or survive some ordeal in order to prove their worth. Male protagonists are encouraged to take the initiative, discover things and solve problems actively.¹³ Female protagonists by contrast find success in passivity and acceptance of their proper place within the family home.¹⁴ Indeed, the home, in contrast to the forest, is generally depicted as the reward for characters that behave in the proper way. The bourgeois idea of home as a place of safety, comprising its own patriarchal moral and legal order, is a neat and easy metaphor for the moral certainties that fairy tales appear to herald. Hansel and Gretel (in Grimm’s tale of the same name) are clearly good children, notwithstanding their greed for sweets, because they are possessed of no greater desire than to find their way home. The story of their unjust expulsion from home by a selfish stepmother is a warning about the need to protect the integrity of home from things that ought to remain outside in the woods. *Hansel and Gretel* is typical of the genre in this respect. The moral culpability of the stepmother is that she has usurped not only the mother’s role, but the father’s as well as master of the house. The threat she poses to the patriarchal order of the home and her wilfulness marks her out as archetypal villain. The world is full of dangers, we are warned, and these must be prevented from entering the home and disrupting its moral order.

Gender and justice are so closely linked, that it is possible to say that characteristics favoured in males would generally constitute wickedness in a female. Sandra Gilbert argues that the options presented to female characters are so narrow as to constitute an imperative. For example in *Snow White*, there is clearly only one viable option for the female protagonist: she must strive become a dutiful wife and mother¹⁵ or else endure the humiliating end suffered by her step mother who did not. As we shall see in detail below, the femininity of fairy tales reflects a conservative attitude towards the moral education of young women. In its immature form, femininity is self-destructive in its naïve curiosity, narcissism and sexuality, and stands in need of correction and protection from its own childish desires. Hence, female characters tend either to be rewarded for accepting the patriarchal order (most female heroines) or else punished for refusing to do so (most female villains). The ordeals necessary for a female character to achieve the prescribed happy ending (a family) have been interpreted as misogynistic. Commentators such as Maria Tatar point out that it is a common characteristic of fairy tales that a female protagonist must endure intolerable

¹² S. S. Jones, *Fairy Tales: The Magic Mirror of Imagination* (Twayne, 1994) p.20

¹³ Compare, for instance, *Tom Thumb* and *The Fisherman and His Wife* from Grimm’s collection. In the former, little Tom is eventually rewarded for managing to survive all manner of troubles (including being eaten by a wolf) where-as as the Fisherman in the other story is punished for giving to his wife’s greedy demands.

¹⁴ Jones, above, at pp.27-8, 65; we shall examine the characters of Sleeping Beauty and Snow White as representatives of such protagonists below.

¹⁵ *ibid.*, at p.66.

abuse and degradation before she can achieve contentment.¹⁶ At least in tales involving a female protagonist, violence against and humiliation of women often appears to be central to the meaning of fairy tale justice. As we shall see, the price paid by Little Red Riding Hood (eaten by a wolf), Snow White (poisoning, choking, prolonged unconsciousness), Bluebeard's wife (attempted beheading) in exchange of their being taught a moral lesson is high indeed.

Fairy tales as moral lessons in obedience, self-restraint and appreciation of family

The story best known as 'Little Red Riding Hood' and related tales, in which a girl's mission to deliver supplies to her Grandmother's house in the woods is interrupted by a hungry wolf, is a story that has received much academic attention in terms of its implications for justice and gender. Each of the two most well-known versions – Perrault's *Little Red Riding Hood* and Grimm's *Little Red Cap* – involve an encounter between a girl and a wolf first in the forest and then at the house once the Grandmother has been eaten. In each version, the girl is shown to be partly responsible for her own and her Grandmother's ordeal, since she allows herself to be diverted from the path and unwittingly directs the wolf to her Grandmother's cottage.¹⁷ The 'moral' at the end of Perrault's version makes it clear that the 'wolf' is really a smooth-talking seducer and that the girl bears at least some responsibility for her demise: "Alas for those girls who've refused the truth, the sweetest tongue has the sharpest tooth." Grimm's version removes most of the evidence of paedophilia, leaving the barest hint of a sexual motive.¹⁸ Both versions are commonly interpreted as a warning about the dangers of immature femininity when faced with dangerous masculinity. For Christina Bacchilega, both Perrault's and Grimm's versions depict femininity as constituted by and secondary to masculinity. For example, Perrault's tale ends with both the girl and the Grandmother actually killed and eaten by the wolf,¹⁹ whilst in Grimm's tale she and her Grandmother must wait to be saved by the honourable woodcutter. In both accounts therefore, the girl is not equipped with the power to determine her own fate and is punished for her weakness. This criticism is also made by Zipes, who argues that the protagonist of Perrault's version especially is "a delicate, bourgeois type, who is helpless, naïve and culpable, if not stupid".²⁰

Bruno Bettelheim gives a more sympathetic interpretation of the story, arguing that, in Grimm's version at least, there is a useful lesson to be learned about harnessing and restraining one's own desires in the interests of duty and safety. Little Red Riding Hood shows her immaturity in being so easily tricked by the wolf. Her subsequent punishment by being eaten is

¹⁶ See S. Sellers, *Myth and Fairy Tale in Contemporary Women's Fiction* (Palgrave, 2001) at p.9.

¹⁷ Grimm, n.1 above at p.134.

¹⁸ For example, the wolf says to the girl: "What are you carrying under your apron?" (Grimm, above, n.1 at p.133).

¹⁹ C. Bacchilega, *Postmodern Fairy Tales: Gender and Narrative Strategies* (University of Philadelphia Press, 1997) at p.58.

²⁰ J. Zipes, *Fairy Tales and the Art of Subversion: The Classical Genre for Children and the Process of Civilisation* (Routledge, 1991) p.29.

right, he argues, because by her own weakness she allowed her Grandmother to be eaten.²¹ The path through the woods is clearly the path of law for this little girl, and her second encounter with the wolf is the enforcement to the law. Although the consequences of her deviance from it are terrible and traumatic, they nonetheless serve only as an admonition. Unlike the archetypal wicked stepmothers who, as we shall discuss, have departed too far from the path to be brought back, Little Red Cap is given a second chance to behave. The experience of being eaten (or whatever else we are supposed to imagine) teaches girls that to practise self-restraint and obey instructions is to protect oneself from the dangers of the world.²²

The importance of conformity in the fairy tale idea of justice is identifiable in many tales involving the ordeals of adolescents. In Grimm's *Snow White*, we are told that a childless queen wished for "a daughter, with skin as white as snow, lips as red as blood, and hair as black as ebony".²³ The wish is granted before the queen dies and is replaced by another woman. As the magic mirror later insists, Snow White becomes "a thousand times more fair" than her stepmother, precipitating the events of the story.²⁴ The significance of the story for our interest in ideas of justice lies in the identity of the authoritative voice of the mirror. Whose voice does the mirror represent? McGlathery argues that it is the voice of the queen's own vanity and narcissism.²⁵ This view certainly makes sense, as it is natural to think of a mirror reflecting one's image and one's own anxieties about it. In this story it illustrates the queen's moral failings as one who has never been able to embrace the joys of family life, which would have brought her to see her own image in the context of her care for others. As such a character, she is incapable of regarding Snow White as her daughter in need of love, but only as a reminder of her own fading beauty. For Bettelheim, it is the voice of Snow White herself who, although initially thinking her mother is the most beautiful woman of all, becomes aware of her own beauty eclipsing her mother's.²⁶ However, it is also possible to regard the mirror as having a constitutive effect upon the characters. Bacchilega argues that, as the voice of truth, the mirror frames, not just the queen's image, but that of femininity generally according to patriarchal lines. "[I]t is a patriarchal frame that takes the two women's beauty as the measure of their (self) worth, and thus defines their relationship as a rivalry".²⁷ It is the narrator of the 'truth' of womanhood: that beauty is everything, and the relationship between women is one of rivalry. I would agree that the voice of the mirror is an important framing device for the story. Bacchilega's analysis is spot-on in showing how the mirror speaks for the women. However, I would suggest that the mirror's presumed neutrality, its role as mediator between characters that it constitutes as rivals, and as a provider of a criteria for judgment means that it

²¹ Bettelheim, B. *The Uses of Enchantment: the meaning and importance of Fairy Tales* (Thames & Hudson, 1976) at p.173.

²² *ibid.*, at p.182.

²³ Grimm, above, n.1 at p.215.

²⁴ *ibid.*, at p.219, 20, 21.

²⁵ J. McGlathery, *Fairy Tale Romance: The Grimms, Basile and Perrault*, (University of Illinois Press, 1991) p.122.

²⁶ Bettelheim, above, n.21, at 207.

²⁷ Bacchilega, above, n. 19, at p.34.

is in fact the voice of law. It determines the criteria of judgment and the relationships between the characters. Enforcement of the law in this case distinguishes between admonition and execution. Firstly Snow White, although showing weakness in allowing herself to be tricked into disobeying the dwarves' injunction on her not to open the door to strangers must endure three apparent deaths and a lengthy spell in a comatose state before she can awaken a new woman. Before this occurs, Snow White herself shows herself to be dangerously narcissistic.²⁸ The reason why the jealous queen so easily overcomes her three times is because Snow White is herself obsessed by her own beauty and therefore cannot resist the items that may make her look even more so.²⁹ However, unlike her stepmother, she is not beyond correction. When the girl is apparently killed by the poisoned apple she has entered the chrysalis-like state common to various other tales necessary in order to be reborn chastened and ready to obey and serve a husband (the prince who finally rescues her).³⁰ Bettelheim imbues his own analysis with an empowering interpretation: his female characters are "twice-born" in order to gain the independence and wisdom to deal with "future wolves".³¹ For Bettelheim, the transformation of girl to woman simply represents the learning curve of adolescence. Of course, a more critical reading would seek to examine the nature of the lesson learned here: is it independence and empowerment or is it merely obedience? On the other hand, the queen, eaten up by her own jealousy can expect nothing but death, and by the cruellest of means.

There are other variations on the Snow White theme that appear in Grimm's collection. *The Almond Tree*, for instance, also involves the creation of a child that, like Snow White, reflects the purity of nature. During wintertime, a childless woman cuts her finger while standing under her Almond Tree. Watching the drops fall on the snow she says: "if only I could have a child as red as blood, and as white as snow".³² The natural environment and childbirth are presented as unified in the shedding of blood and the passing of the months, during which the tree bears its fruit and the woman bears the child she desires. In this tale the child born is a boy with skin as white as snow, and when his mother dies he is brought up by his doting father and jealous step-mother. Unlike Snow White, there is just one (patently successful) attempt by the stepmother to murder the boy when she cuts off his head by slamming a chest-lid down on his neck. She cooks the body for her husband's dinner and afterwards his bones are laid under the Almond Tree. As a confirmation of the irrepressibility of natural life, the magic tree has the boy resurrected as a beautiful bird that finally takes its revenge on the stepmother. We are clearly supposed to believe that the stepmother deserves her fate because she did not respect the natural cycle of life represented by the Almond Tree (that Bacchilega would presumably interpret as patriarchy). The Almond Tree represents all that is natural, whilst the stepmother, like most of her kind in fairy tales, cannot bring herself to leave behind her childish egocentricity and accept the natural order of things. Unlike Snow

²⁸ Bettelheim, above, n.21, at p.203.

²⁹ *ibid.*, at p.209.

³⁰ *ibid.*, at p.213.

³¹ *ibid.*, at p.183

³² Grimm, above, n.1, at p.186

White the boy is simply an innocent, and is allowed to take his own revenge on his stepmother. It is no coincidence that he chooses to do so by dropping a millstone on her, physically reflecting and hence balancing the injustice done to him.³³ The more passive Snow White must content herself by watching the prince's choice of torture and execution of the stepmother queen at their wedding, at which the latter is made to dance on red hot iron shoes until she falls down dead.

The story of Sleeping Beauty and the many other related tales found in European sources forms another example of the inevitability of the reproduction of life, and family life in particular. Typically conservative, Bettelheim reads this tale as being about a father paralysed with anxiety over the unavoidable fact that his little girl will one day grow into a woman and leave him. Despite his meticulous efforts, he cannot prevent the coming of age of his daughter: the pricking of her finger and subsequent bleeding representing the inevitable beginning of womanhood and sexual awakening.³⁴ It is easy to regard this story as it is interpreted by Bettelhiem's view as another affirmation of the status quo: a girl cannot be protected from discovering her true role as a woman in the joys of marriage and children. Grimm's *The Sleeping Beauty* is predated both by Perrault's *The Sleeping Beauty in the Wood* and Basile's *Sun, Moon and Talia*.³⁵ The latter especially contrasts the sheer force of heterosexuality and family life against the futility of the father's efforts to prevent it and the passivity of the female protagonist. In Basile's story, when the heroine (called Talia) 'dies' as prophesised, her heartbroken father-king immediately leaves the kingdom forever.³⁶ As in *Snow White*, feminine beauty is here associated with passivity and surrender since the prince's desire is stirred by the girl's silence and stillness.³⁷ However, whereas in Grimm and Perrault, the princess is awoken by a prince's kiss,³⁸ Basile has a passing king make love to the girl's lifeless body and thereby impregnating her.³⁹ Remaining 'dead' throughout, Talia later gives birth to twins (called Sun and Moon) who eventually bring her back to life by sucking the poisoned hemp from her finger. Whereas Grimm ends with the awakening and marriage, Perrault and Basile both depict the girl's continuing troubles involving her lover's jealous wife (Basile) or mother (Perrault). Predictably, the patient, docile beauty of the young girl is rewarded at the expense of the queen whose beauty is fading and is consumed with murderous jealousy. In Basile's story, the queen demonstrates her resistance to the patriarchal order by inducing the secretary to act contrary to his alliance to the king in helping her to kill Talia and the children she gave birth to whilst asleep. The queen is therefore not only guilty of attempted murder, but also treason. She has what she thinks is the

³³ *ibid.*, p.195.

³⁴ Bettelheim, above, n.21, at p.232-3.

³⁵ Basile, above, n.3 at pp.372 – 376. See also <http://www.pitt.edu/~dash/type0410.html#basile>.

³⁶ Basile, above, n.3 at p.373: “. . . and closing the doors, being desirous to forget all and to drive from his memory his great misfortune, he abandoned for ever the home wherein he had suffered so great a loss.”

³⁷ See Bacchilega, above, n.19, at p.35.

³⁸ Grimm, above, n.1, at p.208.

³⁹ Basile, above, n.3 at p.374: “[He] carried her to a bed, whereon he gathered the first fruits of love. . .”

children cooked and served to her cheating husband. The queen is eventually punished by being burned to death in the fire that she had herself prepared for Talia.⁴⁰

We may presume that the jealous wife of *Sun, Moon and Talia* got what she deserved. The patriarchal order must be respected even if this means accepting that one's husband will not be true to his wife. A 'good' woman maintains quiet dignity in the face of her husband's indiscretions. This does seem to fit with the picture of femininity as synonymous with silence painted in the Snow White and Sleeping Beauty stories. In treating her husband's infidelity as a personal insult worthy of calculated and cruel retribution, Basile's queen has effectively attempted to create a different moral scheme that threatens an existing one. This is unacceptable and so the queen must be eliminated. We should also note that the jealous queen in *Sun, Moon and Talia* has herself not borne the king any children. When she watches her husband eating what she thinks is his children's flesh, she informs him that the meat he is eating is 'his own'. The king's retort is equally suggestive: "I know very well I'm eating what's my own, because you have brought nothing to this house."⁴¹ Thus the appearance of Talia and her two children contrasts against the queen's own failure to do her duty as a woman and a wife.⁴² The bourgeois ideal, of course, is productivity and reproduction of family life in accordance with nature. The queen has not only failed to produce any children of her own, but has jealously and cruelly attempted to kill those of the king. She clearly represents the very opposite to the ideal of family life and is therefore fit for nothing but the flames.

Perrault's *Bluebeard* is one of a number of folk and fairy tales in which the heroine's prize for surviving an ordeal is not marriage, but escape from death at the hands of a powerful and brutal man. Amongst the nastiest of the European tales, they all share the theme of a woman's discovery that a certain man is a misogynistic serial killer and intends to kill her. They all follow one of two possible plot-lines. One form, (including Perrault's *Bluebeard* and Grimm's *Fetcher's Fowl* the woman discovers this through succumbing to the temptation of entering a room from which her husband has forbidden her, and therein discovering the bloody corpses of his previous wives.⁴³ The other form (including the English tale *Mr Fox*, the American *Pretty Polly* and Grimm's *The Robber Bridegroom*) involves a woman making a journey to the house of a man she does not know well (sometimes her suitor or fiancé) where she witnesses him killing a woman in a sexual or cannibalistic ritual. Commentators have suggested that at the very least the stories represent female anxiety about marriage and the consummation of marriage.⁴⁴ Warner identifies fear of childbirth, until the advent of antibiotics and anaesthetics a very real danger for women, as the warning intended for peasant women in the original folktale: "In the forbidden chamber, Bluebeard's wife had perhaps found herself face to face with the

⁴⁰ *ibid.*, at p.376.

⁴¹ *ibid.*, at p. 375.

⁴² Warner, above, n.10 at p. 220.

⁴³ Perrault, above, n.2; Brothers Grimm, *Selected Tales*, trans. D. Luke (Penguin, 2004) pp.362 - 365

⁴⁴ McGlathery, above, n.25 at p. 55.

circumstances of her own death”.⁴⁵ A strong theme in Perrault’s version is that the woman brings disaster upon herself by her own curiosity and inability to obey instructions. It is possible that, as a folk tale that existed long before Perrault’s version was published in 1697, it was told amongst the peasantry, either as depicting the dangers of curiosity and temptation generally, or as a specific response to the case of the French nobleman Gilles de Rais who confessed to the killing of 140 boys near his home in Brittany in 1440. Bettelheim predictably interprets the story as being primarily about the consequences of a woman’s sexual transgression and subsequent guilt. In both Perrault’s *Bluebeard* and Grimm’s *Fetcher’s Fowl*, the woman’s disobedience is discovered because she drops a magic object entrusted to her which becomes indelibly stained with the blood of the previous victims. The tell-tale bloodstains on the key (that Bluebeard entrusts to her and forbids her from using) and on the egg (that the sorcerer compels the women in *Fetcher’s Fowl* to always carry) are the signs of her defloration by another and her guilty conscience.⁴⁶ Christina Bacchilega argues that the woman is no passive victim, but her energy is not directed constructively. Instead she is self-destructively drawn towards a man (Bluebeard) by his riches, power and physical strength despite being repulsed by his appearance and despite knowing that he has been married several times to women, all of whom have mysteriously disappeared. Perrault’s story makes it clear that she suspects the worst by the fact that she instructs her brothers to be ready to rescue her if needs be. In Grimm’s *The Robber Bridegroom* and related stories, the woman travels to and ventures inside her bridegroom’s house despite repeated warnings that by doing so she is putting herself in mortal danger. Significantly, one such warning is issued by the bridegroom himself – “My house is a long way in the wood”⁴⁷ – a warning familiar to fairy tale readers accustomed to the inside/outside distinction denoting safety and danger respectively. In the closely related English folktale *Mr Fox*, the woman ventures into her suitor’s castle despite being warned: “Be bold, be bold, but not too bold, lest your heart’s blood should run cold.” Bacchilega argues that the woman shows herself to be complicit in that which victimises her.⁴⁸ Her adventurousness, curiosity and economic dependence are all directed towards her near-death experience.⁴⁹ Angela Carter’s modern retelling of the story in *The Bloody Chamber* also preserves the theme of female masochism. Her version emphasises the husband’s power over her in every respect, and her acceptance of this. Unnamed as in Perrault, she describes herself as sexually objectified by him, and as one of his jewels on display: “I wore a sinuous shift of white muslin tied with a silk string under the breasts. And everyone stared at me. And at his wedding gift”.⁵⁰ The wedding gift happens to be a “choker of rubies, two inches wide, like an extraordinarily precious slit throat”, surely a macabre warning of what may be to come.⁵¹ On the

⁴⁵ Warner, above, n.10 at p.263.

⁴⁶ Bettelheim, above, n.19 at pp.300-1.

⁴⁷ Grimm, above, n.1 at p.174; the David Luke’s translation for the Penguin edition is even more chilling: “My house is out there in the dark forest.”

⁴⁸ Bacchilega, above, n.17 at p.129.

⁴⁹ *ibid.*, p.138.

⁵⁰ A. Carter, ‘The Bloody Chamber’ in *Burning Your Boats: Selected Short Stories* (Vintage, 1995) pp.111-143 at p.114.

⁵¹ Bacchilega, above n.17 at p.123; Carter (*ibid*) at p.121.

wedding night she is “impaled” by him in love-making. There is surely nothing empowering for femininity in Carter’s prose.

Our review of fairy tales involving female protagonists and subsequent critical literature would suggest that interpretation cannot avoid the intersection between justice and gender. On this reading, justice is served by ensuring that characters that can be educated to accept and value the ideal of productive family life are eventually rewarded. Those that cannot be educated must be eliminated. The ‘justice’ meted out in the fairy tales is a demonstration that the replication of family life is irrepressible. Although politically very different, Bacchilega’s and Bettelheim’s views both attest to the rigidity of the educative moral narrative. All the female protagonists of these fairy tales *necessarily* endure a greater or lesser ordeal at the hands of a cruel villain, suffer a quasi-death, and then are reborn as better, contented adults. The problem with Bettelheim’s analysis is that it implies a meek acceptance of the *status quo*. Bacchilega’s problem is that she can only account for the morality of desert as patriarchal. I suggest that the moral order and thus the fairy tale idea of justice is a broader notion of ‘law’ as a structure of normativity with its own method of enforcement. The natural order implied in the tales indicates that the moral norms are treated as rules (characters either already know or are made to learn them) and these rules are backed by threats of force. The criteria for judgment is objectively determined by the moral order itself and is not open to negotiation by the characters themselves. The voice of law features in all of the tales examined here: we find it in the character of Bluebeard who tries to enforce his will on his wife; in the path through the woods that Little Red Cap must dutifully follow and not veer from; in Snow White learning to be obedient. It is this law that is determinative of a character’s place within the moral hierarchy of fairytale justice. It is sometimes broken – both Little Red Cap and Snow White’s stepmother break the law in their own way – but it is always enforced. Transgressions are punished and the severity of the punishment depends upon the possibility of the offender being brought back within law’s parameters. Where a character is beyond redemption then only the death penalty can be appropriately severe. Law’s ‘others’ must learn to forget their otherness by accepting and obeying the rules. Difference as such is accounted for only by the jealousy, selfishness, narcissism and cruelty of stepmothers and wives, who, being beyond redemption have only humiliation and death to look forward to.

In Search of Justice: Deconstructing the Moral Hierarchy

Our foray into the justice of fairy tales as a moral hierarchy would seem to show a clear distinction between the prioritised ‘inside’ where the safety and domestic order is maintained, and the ‘outside’, the latter being the proper place for wolves and stepmothers. Violence of varying degrees is used by the heroes and heroines of the tales to ensure that justice is appropriately done.⁵² Thus Grimm’s hungry wolf is repaid for his voracious appetite by having his belly filled up with heavy stones. The beauty-obsessed stepmother queen is disposed of by being made to dance in red-hot iron

⁵² M. Tatar, *The Hard Facts of the Grimm’s Fairy Tales* (Princeton University Press, 1987) p.183.

shoes.⁵³ The queen whose heart burns with jealousy for her husband's lover in *Sun, Moon and Talia* is burned in the fire that she prepared for her victim. As Tatar argues, to forgive the villains would leave the stories with a grave moral imbalance.⁵⁴ Indeed, the punishment of wrongdoers may be regarded as a moral imperative. Liberal retributivists argue that the pain suffered by an offender is a thing of value in itself.⁵⁵ If offenders go without punishment then there is no acknowledgment that the offender was responsible for his or her own choices,⁵⁶ and there is no redress for the injury suffered by the community.⁵⁷ Although doubtlessly cruel, it may be a little problematic to accord full responsibility to villains for their actions if we are to accept Bacchilega's insistence of the determinative power of patriarchy.

However, the violent punishments meted out on the fairy tale villains also protect the integrity of patriarchal moral authority. Certain analyses of law, such as that of Peter Fitzpatrick, suggest that law's violence is legitimised by its being proportionate, necessary and in accordance with pre-determined parameters – in opposition to the uncontrolled, 'savagery' of illegitimate violence that knows no rational limits.⁵⁸ In fairy tales, the choice of punishment is a retrospective balance and a restoration of the natural moral order. Since the 'moral' of many fairy tales is that a particular villain brought punishment upon themselves by knowingly transgressing accepted norms, it is arguably proportionate and pre-determined in the sense that they are treated as appropriate responses to the villain's previous wickedness. Thus, fairy tale punishment makes its own appeal to legitimacy. In law, the necessity of force is demonstrated by the need to protect the integrity of the system itself: criminal wrongs between people are also wrongs against the system and these must be punished in order for the system to remain effective. Likewise, the legitimacy of the use of violence against the villains of fairy tales depends upon its necessity in protecting the safety and integrity of the home from outside threat. Justice in this sense requires constant vigilance in defending the border between a conceptual 'inside' (home, safety, order, family, familiar) and outside (danger, disorder, unfamiliar). This section comprises a detailed critical examination of this moral distinction between interiority and exteriority and argues that there is more at play in the narratives than a mundane replication of bourgeois family values. Rather than bolstering this moral integrity, the violence threatened and inflicted in fairy tales actually undermines it.

⁵³ Grimm, above, n.1 at p.224.

⁵⁴ M. Tatar, *The Hard Facts of the Grimm's Fairy Tales* (Princeton University Press, 1987) p.181.

⁵⁵ H. L. A. Hart, *Punishment and Responsibility: Essays in the Philosophy of Law* (Oxford University Press, 1978) p.8.

⁵⁶ A Duff and A von Hirsch, 'Responsibility, Retribution and the 'Voluntary': A Response to Williams' (1997) 56(1) CLJ 103 at 110-112.

⁵⁷ R. A. Duff, 'Punishment, Communication and Community' in M. Matravers (ed) *Punishment and Political Theory* (Hart Publishing, 1999) p.61.

⁵⁸ Fitzpatrick, P. *The Mythology of Modern Law* (London and New York, N.Y., Routledge, 1992) p.81.

Re-reading fairy tale gender stereotypes

The gender roles depicted in the ‘official’ published tales should not be read as indicative of universal truths. The folk tales that had existed in oral form since the Middle Ages were interpreted by Grimm, Perrault and Basile in the context of their respective social conditions. In removing details that they considered unsuitable for their middle class audience, and imbuing the tales with their own cultural values, Grimm and Perrault were both collectors and authors. Zipes describes the Grimms as “bourgeois missionaries” in this respect. Comparing the Grimm’s earliest manuscripts with the 1812 printed edition, Zipes concludes that “each and every oral tale was conscientiously and, at times, drastically changed”.⁵⁹ Viewed as such, we can challenge the gender assumptions of the morally weak woman of, say, the Bluebeard-type tales by applying a different emphasis. Bacchilega does so by insisting that the tale is really all about a woman daring to defy her husband in order to acquire knowledge, and her triumph over death through bravery and mobilisation of family alliances.⁶⁰ Warner argues that other versions of the story (e.g. a nineteenth century story by Italo Calvino called *Silver Nose*) emphasise, not the inability of the woman to resist temptation, but the way in which “quick witted female double dealing overcomes the tyrant”.⁶¹ Thus it is possible to shift the focus from the wife’s indiscretion to her daring escape that is central to Grimm’s *The Robber Bridegroom*. “Female disobedience is hardly the point”, writes Bacchilega.⁶² However, Bacchilega maintains her thesis that the wife’s masochistic, curious and daring character makes her complicit in her ordeal. This makes it difficult to interpret the *Bluebeard* tales. On the one hand it leaves the essentially conservative macho male fantasy that even strong, intelligent women will throw away their own independence and autonomy for a powerful man. However, her cool cunning, determination to acquire riches and above all her drive to gain knowledge suggests not a weak victim but a worthy adversary for her husband. Both male and female characters in the Bluebeard-type tales are preoccupied in attempting to gain power and advantage over the other: Bluebeard and the Sorcerer of *Fetcher’s Fowl* do so by entrusting their respective women with a task requiring a degree of self-restraint that they expect not to be achieved. The women do so by daring to defy a man that regards himself as law.

In *The Robber Bridegroom*, the heroine is so determined to discover the truth about her bridegroom that she ignores all the warnings given to her. In the previous section we suggested that this might represent the masochistic female desire that Grimm and Perrault wanted to show to be in need of tutoring and taming. These tales arguably function to warn women against attempting to acquire knowledge that they are unable to cope with. This argument might carry some weight in the case of Perrault’s *Bluebeard* where the wife may not have been in danger had she obeyed her husband and resisted temptation, but in *The Robber Bridegroom* and related tales there is no such opportunity, since the man kills simply because he is a sadistic

⁵⁹ Zipes, *Fairy Tales and the Art of Subversion*, above, n.20 at p.49.

⁶⁰ Bacchilega, above, n.17 at p.107-9.

⁶¹ Warner, above, n.10, p.257.

⁶² Bacchilega, above, n.17 at p.111.

misogynist. It is vitally important for her own survival that the woman learns what he is really like. Her perseverance does not come from a self-destructive will then, but from bravery and strength of character. We would also have to concede that the function of the key (*Bluebeard*) and egg (*Fetcher's Fowl*) is undecidable in terms of determining the just deserts of the female protagonists. Unlike Oedipus, for whom discovery of the whole truth brought inevitable disaster, the female protagonist in these tales is rewarded for her troubles by retaining the husband's riches, as she always desired. Her apparent character flaws of failure to resist temptation and attraction to danger are also her character strengths, as they are necessary for her to learn about the world in which she lives, and hence for her acquisition of power.

Both the villains and heroes of the Little Red Riding Hood motif can be deconstructed by considering the tale in its specific, historical context. For instance, a traditional Italian folk tale (written down in Italo Calvino's *The False Grandmother*⁶³) has the girl being tricked by the wolf into eating her Grandmother's boiled teeth ("beans") and fried ears ("fritters"), stripping and getting into bed with him. Perrault and Grimm also altered the scope of the gender roles. Calvino's version credits the girl with far more intelligence and resourcefulness than did Grimm or Perrault. For instance, the famous exchange between girl and wolf girl on the 'Grandmother's' strange physical appearance (which is common to all versions of the tale) leads, not to her being eaten, but to her concluding that, being so physically strange ("why is your chest so hairy, Grandmother?"), the thing in the bed cannot be her Grandmother. Calvino then has the heroine excusing herself to defecate in the barn below the cottage, before making her escape and then finally killing the beast as it tries to pursue her. Nor does Calvino suggest that the girl is herself to blame for the events. His version lacks the initial encounter in the forest in which the girl implicates herself in the wolf's crime by directing him to her Grandmother's house. Instead, her journey to her Grandmother's house has her making peace offerings to a gate and a river from her basket, which in return later help her to escape. Perrault's decision to have her finally eaten and Grimm's to have her rescued by the woodcutter significantly alters both the tale's theme from a brave girl's adventure to moral correction, and its view of the proper level of activity for a female character.

Nor is the Grandmother all that she seems. The Grandmother of Little Red Riding Hood is supposed to represent the end point of the mission that the girl's Mother sends her on. But instead of finding the safety of home from home, she finds the wolf waiting to eat her up. As we have discussed, the Red Riding Hood of Grimm and Perrault has unwittingly directed the wolf there and the Grandmother invited him in. However, is it possible that the wolf was already living the cottage before the story's events unfold? Whereas in Grimm and Perrault, the wolf and Grandmother are clearly different characters, Angela Carter's story *The Werewolf* suggests that the wolf is actually part of the family, specifically, that the wolf *is* the Grandmother: that she is a Werewolf.⁶⁴ In Carter's story, instead of straying from the path

⁶³ <http://mld.ursinus.edu/Maerchen/false.html>.

⁶⁴ Carter, above, n.50, p.210 – 211.

and naively directing the wolf to deal with her Grandmother, the girl bravely and competently fights it off with a hunting knife, cutting off its front paw. When she gets to the cottage she discovers her Grandmother's true identity by the tell-tale bloody stump where a hand should be and duly has the werewolf killed. Angela Carter's story alludes to the historical fact that remote peasant communities of the late Middle Ages Europe lived in fear of wolves, and that popular association between wolves and the devil sustained a belief in werewolves. Records indicate that it was not uncommon in the sixteenth and seventeenth centuries for an accusation of being a werewolf to be used as a form of scapegoating.⁶⁵ Can we not read Little Red-cap's own perception of wolf-like qualities in her Grandmother ("what large hands you have got!"⁶⁶) as a dawning awareness that 'family' does not always mean 'safety'? It would be mere comedy if the girl were actually to confuse her own grandmother with a wolf wearing her nightcap and glasses. The literary critics of the 1970s wondered why the Grandmother is so easily tricked into letting the wolf into the house when she ought to know better; why she dressed the girl up coquettishly in a pretty cap (for it was a present from the Grandmother); why, further, the mother sends her out into the forest so ill prepared to deal with its dangers. Those critics put it down to Oedipal anxieties and the dangerous naivety of female sexual curiosity and masochism. Zipes and others dismiss such psychoanalysis as "banal, superficial observations" about sexual desire, violence and power⁶⁷ which reduce the tales to "material for pop psychology self-help books".⁶⁸ In the light of Carter's version the guilty sexual desires analysis is not necessary. *Little Red Riding Hood* is a tale about family life and the fears of peasant communities. In this light, the black comedy of the girl's examination of Grandmother's physical peculiarities becomes a horrific lesson that one's own family might not necessarily have one's best interests at heart.

Some commentators argue that this way of reading *Little Red Riding Hood* as a warning about violence in the home should be applied also to the stepmother character in other tales. It is Zipes's view that the wickedness of the stepmother may in fact represent, not the opposite of normal family life, but the family as itself a scene of violence and abuse.⁶⁹ Zipes argues that the Grimms wanted to warn of the abuse of children within the family home. The woman who abandons, neglects, abuses or attempts to murder the children of a family is in fact the children's real mother. McGlathery supports this, pointing out that in the original Grimm version of *Snow White*, the jealous woman is Snow White's mother.⁷⁰ The 'step' mother is thus superimposed by Grimm onto the simpler nuclear family as a literary device for allowing the abuse to be spoken of without utterly destroying the moral integrity of family.⁷¹ Therefore, for Zipes, the stepmother character stands for the dangers that children face within the family. Zipes's view carries the troubling implication that the innocent, beautiful adolescent of the fairy tales

⁶⁵ Zipes, *Fair Tales and the Art of Subversion*, above, n.20, at pp.28-9.

⁶⁶ Grimm, above, n.1 at p.135.

⁶⁷ Zipes, *The Brothers Grimm*, above, n.4 at p.117.

⁶⁸ *ibid.*, p.118.

⁶⁹ *ibid.*, p.120-7.

⁷⁰ McGlathery, above, n.25 at p.122.

⁷¹ Zipes, above, n.4 at p.121.

may naturally mature to become the frustrated, narcissistic mother that turns on her own daughter in envy. The 'norm' may therefore not be the 'happily ever after' of lifelong, happy marriage and childrearing, but an arena of cruel torture and death. Thus, we might discern a twist to the generally assumed static prioritisation of family over egocentric values. It is the family itself that is shown to be the site of violence, but this is too horrible to be narrated without the dissociative artifice of the 'step' mother.⁷²

Thus, the issue being raised in these tales is not the outside threat that the 'step' mother poses, but the threat faced by children already within the family. However, as Marina Warner argues, it is also possible that the stories reflect inter-generational strife within the family. Warner argues that the peasant folk tales that formed the sources for Grimm's and Perrault's official texts had survived since Middle Ages Europe through oral transmission by older women, who narrated the conflicts between younger and older generations. Oral folk tales served as a reminder to the youth of the continued power of the elders over younger family members who may be tempted to marginalise or even desert them. Marina Warner describes how, until the nineteenth century, the English word for step mother carried the same meaning as mother-in-law, and both labels carried the semi-sinister implication that the woman was not a mother by *nature*.⁷³ The cruelty of the fairy tale 'step' mothers towards step-daughters or daughters-in-law is to be associated with the older women's anxieties about becoming dispossessed after the death of their husband and redundancy in society.⁷⁴ In the context of a society in which widows were expected to be cared for by their son's family, resulting in large scale destitution of older women who may have never married, the tales as passed on orally by women serve to warn of the "bitter conflicts within the romance of marriage"⁷⁵ She acknowledges that the quasi-feminist depictions of older women in the oral folk tales did not survive the transition from the peasantry to their eventual 'official' form in Grimm and Perrault. The latter turned the historically specific accounts of women's inter-generational rivalry into a general demonisation and an antithesis to the feminine ideal.

Other dangers in the home

A related point is that the *Sleeping Beauty*-type tales that feature a father-daughter relationship depict not simply the already-discussed fatherly anxiety regarding his daughter becoming a woman, but in also the threat to family stability by the father's incestuous desire for her. Bettelheim is typically alert to the possibility that the father-king that leaves the kingdom and his daughter forever and lover-king that discovers her in Basile's story could in fact be the same person, although for Bettelheim this merely reflects the childish Oedipal desire of girls to seduce their father.⁷⁶ However, such a desire on the part of the girl is manageable within the moral context of the

⁷² Bettelheim remarks that this dissociative device is used by children who invent an imaginary 'step'mother in order to cope with the fact that his or her own mother may not always behave in a loving or caring manner.

⁷³ Warner, above, n.10 at 219.

⁷⁴ *ibid.*, p.229.

⁷⁵ *ibid.*

⁷⁶ Bettelheim, above, n.21 at p.228.

fairy tale because it is just a complex to be worked through on the road to maturity. McGlathery's suggestion that the central theme is not the girl's but the father's desire is altogether more problematic because it threatens to destroy the family as a place of safety. Why does the father-king of Basile's story disappear from the narrative so suddenly when his daughter 'dies' as prophesised? McGlathery suggests that it is because he is afraid that he will be tempted to violate her body.⁷⁷ We might also wonder why the father-king is so anxious to prevent the onset of the inevitable in the first place. If we accept Bettelheim's and McGlathery's interpretations of the father's anxieties to prevent to loss of his daughter to womanhood then it is arguable that the distinction between father and lover is blurred to the point of undecidability. By the father's incestuous desire and the source of violence against children being locatable within the family itself, the status of the family as a place of order and harmony becomes more problematic still.

Given the apparent preoccupation in tales of the 'Snow White' and 'Sleeping Beauty' type with the inevitability of replication and productivity of wholesome family life, it should perhaps not be surprising that we find incest represented. The link between the natural cycle of life and death and incestuous desire lies in male anxieties about loss: a man whose wife dies young is physically reminded of her by their child. At the same time he is anxious that the daughter borne by his deceased wife is fast growing up and will one day leave him. Interpreters of the fairy tales have tended to suggest that the memory of losing the wife coupled with the anxiety about imminently losing the daughter to womanhood and marriage motivates a sexual desire in the father for his adolescent daughter. We have seen how the 'birth' of Snow White is precipitated by sexual desire ("skin as white as snow, a mouth as red as blood") and accompanied by the death of the mother. The drops of blood that the queen watch falling onto the snow are representative both of sexual reproduction and of Snow White's physical appearance (and in being called 'Snow White', this is constitutive of her very identity). It is not surprising, then, that Angela Carter's modern retelling has the newly created snow-child raped by the Count at the end of the story.⁷⁸ In Basile's *Sun, Moon and Talia*, the mother of the girl is simply absent, presumably having died in childbirth. The absence of a mother makes the daughter the only representative of femininity in the tale, and thus it is understandable that the king is, first, so anxious not to let her go, and second, so heartbroken at his failure to prevent the inevitable.

McGlathery argues that the relationship between the king and the princess makes the 'Sleeping Beauty' type tales comparable to those in which the incestuous desire of a father for his daughter is made explicit. Tales in all three major collections involve fathers who desire their daughters because a dying wife implicitly condones or even requests it. Grimm's *Cat-Skin* and Perrault's equivalent *The Donkey's Skin* both involve a dying queen making her husband-king promise never to remarry unless he finds someone quite as beautiful as she.⁷⁹ In both the tales, the initially inconsolable king eventually

⁷⁷ McGlathery, above, n.25 at p.99.

⁷⁸ Carter, above, n.50, pp.193-194, 'The Snow Child'.

⁷⁹ See The Brothers Grimm, *Household Tales* (Dent & Sons, 1953) pp.215 – 220 at p.215.

realises that the only woman who can match his dead wife's beauty is his own daughter, and ignoring dire warnings from his courtiers, tries desperately to have her. Both tales show the daughter driven into squalid hiding in order to escape her father's amorous advances. Disguised in rags, she becomes a servant at the palace where she coyly allows herself to be discovered by a Cinderella-type plot device (involving a gold ring that she drops into the king's soup) and readmitted into royal society. Perrault's tale makes more moral sense than Grimm's, ending with the girl's marriage to a prince and her forgiving her repentant father. Incredibly, Grimm's tale ends 'happily' with the princess entering fully into the role of her father's wife. The delighted king, who seems to have forgotten the identity of this girl, exclaims: "You are my beloved bride, and we will never more be parted from each other."⁸⁰ Is it possible that the sleeping beauty theme is related to these tales of fatherly incestuous desire? Furthermore, can justice for the protagonists of these fairy tales really incorporate necessary exposure to incest and abuse? If we were to accept that bad things befall good people in order that they can mature then we could say yes – it is still just that they are 'rewarded' with the flawed prize of such domestic ordeals. If, as McGlathery argues, Sleeping Beauty is to be associated with the incest tales, then the moral significance of these tales is that the prize is not pure or innocent and the family is no longer simply the site of productivity and new life.

Conclusion

However we choose to interpret the tales examined here, the only certain thing is that the juxtaposition of safety to danger, good to bad, aggressor to victim, etc, is not a clear distinction. Of course, the destabilising of hierarchical oppositions means that a simple definition of 'justice' as it is presented through the medium of fairy tales actually seems further away. The ideal of the home that is defended against threats from outside is undermined by the ambiguities of characterisation. Principally, the distinction between inside and outside as indicators of moral order (and thus of justice) come under threat from within. How can one shut the door of the family home against the dangers of the forest when those dangers are always already inside the house? The protective measures displayed in the fairy tale narratives are deployed to do just that, but since the danger is already inside, such measures are reduced to mere covers for the source of the threat. Thus a mother who feels murderous jealousy because of her daughter's superior beauty is distanced from the family by being cast as a stepmother. The wolfish traits of one's own family members are disguised in the narrative by creating the character of the wolf and locating him outside in the forest.

⁸⁰ *ibid.*, p.220.

**PSYCHIATRIC HARM AND THE INVOLUNTARY PARTICIPANT:
“A STORY OF THE EBB AND FLOW OF TORT LIABILITY”¹**

(Case commentary: From *Dooley v Cammell Laird & Co Ltd and Mersey Insulation Company Ltd* through to *W v Essex County Council* and *Salter v UB Frozen and Chilled Foods Ltd*.)

*Vivien Pickford, Senior Lecturer in Law, Portsmouth Business School, University of Portsmouth**

Introduction

Claimants seeking damages in an area of law which has been variously described as “a patchwork quilt of distinctions which are difficult to justify”² and one where “the common law has taken a wrong turn”³ face particular challenges. Certain categories of claimant who had been treated more favourably than others appeared to lose their special status following the House of Lords’ decision in *White v Chief Constable of South Yorkshire Police*⁴ but more recent decisions suggest that this may not be so with regard to litigants claiming as involuntary participants. This commentary traces the development of the law with regard to such claimants and considers the impact of recent cases.

Recognising the Category of Involuntary Participant

The civil courts have always been presented with difficulties when determining claims in negligence for pure psychiatric harm, battling in particular with the “floodgates argument” and a complex, progressing area of medicine. Development of the law has therefore been slow and initially a claimant needed to be able to prove that he had been in reasonable fear for his own safety in order to succeed.⁵ *Hambrook v Stokes Bros*⁶ represented the first successful claim by a secondary victim, a mother who was in no physical danger herself but who reasonably feared for the safety of her children. She was not categorised as such in the judgments and neither was there any attempt to categorise the claimant in *Dooley v Cammell Laird & Co Ltd and Mersey Insulation Company Ltd*,⁷ which is generally

* I am grateful to Professor Maniruzzaman and to Adele Sinclair, at Portsmouth Business School, for their helpful comments on an earlier draft of this article.

¹ Taken from *White v Chief Constable of South Yorkshire Police* [1999] 2 AC 455, at 503 *per* Lord Hoffman

² *ibid.*, at 500 *per* Lord Steyn.

³ Law Commission No. 249 *Liability for Psychiatric Illness* para.4.2.

⁴ *Op cit.* n.1.

⁵ *Dulieu v White & Sons* [1901] 2 KB 669.

⁶ [1925] 1 KB 141.

⁷ [1951] 1 Lloyd’s Rep 271.

acknowledged to be the first case involving a successful claim by an involuntary participant. Dooley was a crane-driver who suffered psychiatric harm when the defective rope on his crane broke, causing the load to crash into the hold of a ship where his fellow employees were working. The decisions in both cases were based on psychiatric harm being reasonably foreseeable.⁸

The distinction between different categories of claimant was first highlighted by Lord Oliver in *Alcock v Chief Constable of South Yorkshire Police*,⁹ a case arising out of the Hillsborough disaster. Unlike previous cases, there were potentially a large number of claimants so that “floodgates” was a particular issue. Lord Oliver famously stated that broadly claimants could be divided into two categories; “those cases in which the injured plaintiff was involved, either mediately or immediately, as a participant, and those in which the plaintiff was no more than the passive and unwilling witness of injury caused to others.”¹⁰

Lord Oliver’s examples of those falling into the first of these categories and therefore a primary victim or equivalent included rescuers and those where:

“. . . the negligent act of the defendant has put the plaintiff in the position of being, or of thinking that he is about to be or has been, the involuntary cause of another’s death or injury and the illness complained of stems from the shock to the plaintiff of the consciousness of this supposed fact.”¹¹

He thereby provided the involuntary participant with special status and gave *Dooley* as one of his examples. In *Dooley* there was in fact no indication that the claimant blamed himself for what happened and neither was there a primary victim other than Mr. Dooley himself as fortuitously none of his colleagues was injured.¹² Lord Oliver, however, also opined that:

“The fact that the defendant’s negligent conduct has foreseeably put the plaintiff in the position of being an unwilling participant in the event establishes of itself a sufficiently proximate relationship between them and the principal question is whether, in the circumstances, injury of that type to that plaintiff was or was not reasonably foreseeable.”¹³

⁸ Dooley recovered damages in negligence against the owner of the defective rope but also recovered damages for breach of statutory duty against his employers, Cammell Laird.

⁹ [1992] AC 310.

¹⁰ *ibid.*, at 407.

¹¹ *ibid.*, at 408.

¹² This is in contrast to his Lordship’s other examples of *Galt v British Railways Board* (1983) 133 NLJ 870 and *Wigg v British Railways Board* The Times, 4 February 1986, unreported, both involving train drivers who did wrongly think that they were responsible for killing or injuring persons. The brief account of *Wigg* additionally uses language appropriate to a rescuer trying to help the primary victim and so it could be argued that the claimant fell into that special category also afforded special status by Lord Oliver.

¹³ *Op cit.* n.9, at 408.

Mr. Dooley was certainly an “unwilling participant” and Lord Oliver’s judgment emphasised the significance of being such a claimant in that the proximity requirement was satisfied. In contrast “a passive and unwilling witness”¹⁴ or secondary victim was and still is faced with a number of control factors in order to establish a duty of care owed to him, one of which is proximity specifically in terms of closeness of relationship to the victim, closeness in time and space to the incident and the means by which the shock was caused.¹⁵ It is the first element of this proximity test which is the most likely stumbling block for an involuntary participant if he were not to be classified as primary or of equivalent status.

Subsequent cases such as *Robertson and Rough v Forth Road Bridge Joint Board*¹⁶ recognised the position of the involuntary participant and distinguished him from a bystander who “was not directly involved in it as the actor by whose hand the death or injury was caused to the third party”.¹⁷ Perhaps the words “or the reasonably imagined death or injury” should have been added in view of Lord Oliver’s examples. Further, the emphasis then appeared to be on the feeling of responsibility for what occurred, despite this not being the case in *Dooley* itself. The pursuers in *Robertson* therefore failed in their claim. Although they and their fellow employee, who was killed when he was blown off the Forth bridge, had all been engaged on the same task in the course of their employment at the time of the incident, it was natural forces which had caused it and they were aware of that.

Arguably more difficult to understand is the decision in *Hunter v British Coal Corporation*.¹⁸ The claimant was driving a truck underground and hit a water hydrant which was protruding into the roadway through the defendant’s negligence. He and a colleague unsuccessfully attempted to stop the flow of water and so he then searched for a hosepipe to divert the flow. Whilst he was some 30 metres away he heard the hydrant explode and then about ten minutes later a message was given out to the effect that a man had been injured. As he walked back towards the scene he was told that his colleague was dead. His claim for damages for psychiatric harm was unsuccessful; it was held that he was not an involuntary participant as he was not present at the scene of the accident. Further his subsequent illness was described as an abnormal reaction, caused by an irrational feeling of responsibility and therefore unforeseeable. Lord Hobhouse, however, who dissented, found Lord Hope’s judgment referred to above in *Robertson* to be the most “illuminating contribution”¹⁹ to the question of whether in the circumstances psychiatric harm to the claimant was reasonably foreseeable. Concern was expressed by the majority that the law should not provide compensation for a survivor’s guilt but it would seem that it was not simply because he survived, and his colleague did not, that Hunter suffered his illness but because he genuinely felt responsible for his colleague’s death as he had hit the hydrant. The defendants conceded that if Hunter had

¹⁴ *ibid.*, at 407 *per* Lord Oliver.

¹⁵ Other control factors include the shock requirement and that psychiatric harm be reasonably foreseeable in a person of normal fortitude.

¹⁶ [1995] IRLR 251.

¹⁷ *ibid.*, at 255 *per* Lord Hope.

¹⁸ [1999] QB 140.

¹⁹ *ibid.*, at 165.

witnessed his colleague being killed then they would have been liable. However, *Dooley* does not and cannot require that the claimant actually sees the death or injury to the third party as no-one else suffered any harm in that case. Mullany questions “On what sound basis can the coal miner worker be denied recovery for proven psychiatric injury sustained on learning of the death of a colleague on site with whom he was working shortly beforehand?”²⁰ The answer must surely be that there is none.

In *Hunter* Lord Hobhouse made reference to the Law Commission Consultation Paper²¹ and its Report on Liability for Psychiatric Illness²² was published shortly afterwards. It recommended that the involuntary participant should be entitled to recover, acknowledging that the floodgates objection would not apply in such cases and that medical literature supported such inclusion.²³ It considered that there should be a special rule as set out by Lord Oliver applicable to involuntary participants and saw no reason to restrict their existing rights.²⁴ Of particular interest is the use of the phrase “special rule” indicating that such claimants were regarded as being in a category of their own and this is supported by the Commission’s recommendation that:

“Although a legislative provision on this would not be appropriate, we tend to the view that the courts should abandon attaching practical significance, in psychiatric illness cases, to whether the plaintiff may be described as a primary or a secondary victim”.²⁵

Regrettably this has not happened; instead the categorisation of primary and secondary victims has become crucial to the determination of liability.

Muddying the Waters

The decision of the House of Lords in *Page v Smith*,²⁶ a case involving a participant in an accident who suffered only psychiatric harm, did not on the facts appear to be particularly relevant in terms of distinguishing between primary and secondary victims. The claimant was obviously primary being directly involved in the accident. However the repercussions of this decision are considerable. It made the distinction between primary and secondary vital, determining that a primary victim does not have to show that psychiatric illness was foreseeable in a person of normal fortitude; reasonable foreseeability of personal injury is sufficient. Thus it “dethroned foreseeability of psychiatric injury from its central position as the unifying feature of this branch of the law”²⁷ attracting considerable criticism.²⁸ It also

²⁰ “English Psychiatric Injury Law - Chronically Depressing” (1999) 115 *LQR* 30 at 31.

²¹ Consultation Paper No. 137, *Liability for Psychiatric Illness*.

²² *Op cit.* n.3.

²³ *ibid.*, para.7.6.

²⁴ *ibid.*, para.7.5.

²⁵ *ibid.*, para.5.54.

²⁶ [1996] 1 AC 155.

²⁷ *Op cit.* n.1 at 475 *per* Lord Goff.

blurred that vital differentiation. Lord Lloyd, for example, described Lord Oliver's distinction as being "the most convenient and appropriate terminology"²⁹ but then adopted a narrower classification by describing a secondary victim as one who is "almost always outside the area of physical impact".³⁰ Other judges in *Page* used alternative terminology such as participant and bystander,³¹ and participant and non-participant,³² and there was no discussion, for example, of *Dooley*.

The necessity for clarity was provided in *White v Chief Constable of South Yorkshire Police*,³³ another complex House of Lords' decision, which like *Alcock* arose out of the Hillsborough disaster. It was, however, arguably to bring the law full circle for claimants such as rescuers and involuntary participants. Emphasis was once again placed on physical injury or reasonable fear therefore and *White* may therefore be seen as a backward step in the development of the law in this area.

Having allowed none of the friends or relatives of the victims to succeed in their claims for psychiatric harm in *Alcock*, their Lordships were reluctant to find in favour of the police officers claiming in *White*. They had succeeded in the Court of Appeal³⁴ on the basis that they were "quasi-employees"³⁵ of the tortfeasor, or rescuers, or both and therefore primary victims. The majority of their Lordships, however, considered employee witnesses to be in no better position than other bystanders.³⁶ Further they determined that rescuers, as such, were not primary victims. This was despite such claimants, like involuntary participants, having been identified by Lord Oliver as examples of primary victims, and despite the supportive conclusions of the Law Commission. The majority interpreted *Page* as restricting the definition of primary victim to those who were either in danger of physical injury or who reasonably believed themselves to be so endangered.³⁷ As none of the claimants could satisfy that definition or, as secondary victims, prove a closeness of relationship to those physically injured or killed, they all failed in their claims.

It can be strongly argued that this is a misinterpretation of *Page* and that whilst it provides us with an example of a victim who is primary, because he was actually involved in the crash and in reasonable fear of being injured, it does not restrict the primary category to such claimants. Being within the range of foreseeable physical injury is, as Cooke comments, "a sufficient condition of liability for psychiatric injury but not a necessary condition of

²⁸ See, e.g. Trindade, "Nervous Shock and Negligent Conduct" (1996) 112 *LQR* 22, Teff, "Liability for Negligently Inflicted Psychiatric Harm: Justifications and Boundaries" (1998) 57(1) *CLJ* 91.

²⁹ *Op cit.* n.26 at 184.

³⁰ *ibid.*, at 187.

³¹ *ibid.*, at 178 per Lord Jauncey.

³² *ibid.*, at 182 per Lord Browne-Wilkinson.

³³ *Op cit.* n.1.

³⁴ *Frost v Chief Constable of South Yorkshire* [1998] QB 254.

³⁵ Police officers are holders of an office rather than employees.

³⁶ This may be contrasted with cases where the employee suffers direct psychiatric harm such as *Barber v Somerset County Council* [2004] 2 All ER 385.

³⁷ Lord Goff dissented on this issue.

such liability”.³⁸ Indeed Lord Lloyd himself in the leading judgment in *Page* concluded with a number of propositions but made no reference therein to determining who is a primary victim.³⁹

It was not, in any event, necessary on the facts of *Page* to answer that question; the distinction was only relevant in deciding the effect of such categorisation. *White* however is very much a policy decision⁴⁰ and signalled bad news in particular for claimants such as rescuers and involuntary participants. It seemed to leave them in an impossible position; having to prove physical danger or reasonable fear thereof, in which case they would simply be primary victims anyway, or a close relationship with the victim which would be most unlikely. In effect the decision appeared to destroy such special categories of claimants.

Turning the Tide

This area of law has been dominated by the Hillsborough litigation and its policy decisions placing the judiciary in a difficult position when faced with claims for psychiatric harm arising out of very different circumstances. Maybe this was foreseen by Lord Hoffman in *White* when he made reference to *Dooley*, *Galt* and *Wigg* and commented that “there may be grounds for treating such a rare category of case as exceptional and exempt from the *Alcock* control mechanisms”.⁴¹ Such a case came before the House of Lords in 2000. In *W v Essex County Council*⁴² parents claimed inter alia damages for psychiatric harm caused by the negligence of the council and its social worker. The claimants had four young children and were approved as specialist adolescent foster carers by the council. They made it clear that they would not accept a child who was a known or suspected sexual abuser but such a boy was placed with them. The boy abused the claimants’ children and both the claimants and their children suffered psychiatric harm. The children’s claims proceeded but those of the parents were struck out by the trial judge,⁴³ a decision which was upheld by the Court of Appeal,⁴⁴ on the basis that they were secondary victims and could not satisfy the *Alcock* control factors. Their appeal to the House of Lords succeeded however.

One of the arguments advanced by the parents was that they felt “they had participated in or contributed to or laid the foundation for the commission of the acts of abuse on their children”⁴⁵ by arranging for the boy to be brought into their home and that they were therefore primary victims. In the leading judgment Lord Slynn recalled “the need for flexibility in dealing with new situations not clearly covered by existing decisions”,⁴⁶ which had been

³⁸ “Primary Victims: The End of the Road?” (2004) 25 *Liverpool Law Review* 29 at 35.

³⁹ *Op cit.* n.26 at 197.

⁴⁰ Policy motives are freely discussed by both Lord Steyn at 498 and Lord Hoffman at 510.

⁴¹ *Op cit.* n.1 at 508.

⁴² [2001] 2 AC 592.

⁴³ [1997] 2 FLR 535, Hooper J.

⁴⁴ [1999] Fam 90.

⁴⁵ *Op cit.* n.42 at 599 *per* Lord Slynn.

⁴⁶ *ibid.*, at 600.

recognised by Lord Scarman in *McLoughlin v O'Brian*.⁴⁷ More surprisingly he considered “the categorisation of those claiming to be included as primary or secondary victims is not as I read the cases finally closed. It is a concept still to be developed in different factual situations.”⁴⁸ He further stated that none of the relevant cases to which he had been referred conclusively showed that if the parents’ injury flowed from a feeling of responsibility they were prevented from being primary victims and cited Lord Oliver in *Alcock*.⁴⁹ He therefore allowed the appeal and Lord Steyn, one of the majority in *White*, and Lord Hope both concurred.⁵⁰

Perhaps therefore the door had always been left slightly ajar for claims by involuntary participants but *W* makes their position stronger and was relied on in the recent Scottish case of *Salter v UB Frozen and Chilled Foods Ltd*.⁵¹ The pursuer was employed by the defenders in its factory as a forklift truck driver and during stocktaking he lifted his workmates in a cage mounted onto the truck to check the stock on the rows of pallets, some of which were close to the roof structure of the warehouse. They wore helmets and at the highest level had to duck from time to time to avoid the beams. Tragically, one of the stock-takers was fatally injured when he hit his head on a cross-member. The pursuer, who was unable to see his workmates when they were working at that level, was aware that his truck had shuddered at one point. He then saw a helmet fall to the ground and as he moved the truck, on the instructions of the other stock-taker, he saw blood dripping from the cage and then the body of the stock-taker, bleeding from various orifices. He blamed himself for the accident and was subsequently diagnosed as suffering from post traumatic stress disorder and depression.

The defenders sought dismissal of the pursuer’s claim for damages in negligence arguing that he was not a primary victim as he was not in physical danger or reasonable fear thereof and he did not satisfy the secondary control mechanisms.⁵² The trial judge, however, held the pursuer to be a primary victim because he was actively involved in the accident, although not himself to blame; the real cause was the unsafe working practices of the defenders. The judge commented that it was not expressed in any of the authorities cited to him that *Dooley* had been wrongly decided; the facts were similar and perhaps stronger as Mr. Dooley had not blamed himself, and it was “at the very least arguable that, having regard to the authorities and in particular *Dooley* and the subsequent treatment of that case, psychiatric injury was, on the assumed facts, reasonably foreseeable”.⁵³

Arguably *Salter* does not sit comfortably with *White* but neither do various obiter dicta statements by members of the House of Lords referred to above. The fundamental problem in this area of law is the requirement to categorise claimants as either primary or secondary and the way in which that classification was restricted in *White*. Whilst the desire to prevent liability

⁴⁷ [1983] 1 AC 410 at 430.

⁴⁸ *Op cit.* n.42 at 601.

⁴⁹ *ibid.*

⁵⁰ The action was finally settled out of court.

⁵¹ 2003 SLT 1011.

⁵² The defenders also argued that the action was time barred.

⁵³ *Op cit.* n.51 at 1019.

extending too widely in cases involving pure psychiatric harm may be understandable, a “one size fits all” approach is particularly inappropriate in cases like *Dooley* and *Salter* where there is no fear of “floodgates” anyway. Further, it has arguably produced some unjust decisions even where there is such concern, as demonstrated by *Alcock* itself. It has already been noted that the Law Commission recommended amendments to the law, such as abandoning attaching practical significance to such classification⁵⁴ and since *Salter* the Scottish Law Commission has also reported,⁵⁵ proposing various changes including the abolition of the “classification as such of persons into primary and secondary victims”.⁵⁶

Its draft bill distinguishes between those suffering mental harm because of what directly happened to them and others who were not directly involved. It is made clear in the explanatory notes that the former would include not only those at risk of physical harm but also “persons who were involved in such a way as to feel personally responsible for it, even though they were not at fault”.⁵⁷ The defender would, however, only be liable to such persons if the harm amounted to a medically recognised mental disorder and it was reasonably foreseeable that the act was likely to cause a person in the victim’s position to suffer such harm. Additional criteria would have to be satisfied by those not directly involved, unless a rescuer.⁵⁸ The proposed legislation would therefore place a pursuer like Mr Salter in the first category but what about someone like Mr. Dooley who did not argue that he felt personally liable? He too should be able to argue that he falls into that category as he was directly involved. The recommendations of neither Law Commission go as far as advocating a single test of reasonable foreseeability but the decision in *Salter* caused Kinloch to question:

“. . . whether the complex primary and secondary victims categorisation with which the courts have to struggle really represents any advance in analysis over the simple reasonable foreseeability test applied all those years ago in the first of the important nervous shock cases to arise in Scotland, *Bourhill v Young*, 1942 SC (HL) 78”.⁵⁹

It cannot be denied that such categorisation has caused more problems than it has solved in cases where “floodgates” is not an issue. However the courts would struggle with a single reasonable foreseeability test where there are, potentially, many claimants. In such cases they are still awaiting “a coherent, logical and just set of principles”.⁶⁰

Conclusions

Members of the judiciary, of the Law Commissions for both England and Wales and Scotland, and academics continue to be critical of this area of law and in particular the division of claimants into primary and secondary

⁵⁴ *Op cit.* n.25.

⁵⁵ Scottish Law Commission No.196, *Report on Liability for Psychiatric Illness*.

⁵⁶ *ibid.*, para.3.3.

⁵⁷ *ibid.*, at 52.

⁵⁸ *ibid.*, *Reparation for Mental Harm (Scotland) Bill [Draft]* s.4.

⁵⁹ “*Salter v UB Frozen and Chilled Foods Ltd*” (2003) 33 *SLTA* 261 at 262.

⁶⁰ *Op cit.* n.29, Trindade at 27.

victims. Further, whilst some would argue that at least the boundary between them was clarified in *White*, others consider the distinction to be far from clear. Many comment too that the law is still evolving and particular examples of this include not only issues such as what constitutes “immediate aftermath”⁶¹ but also, of particular relevance to this commentary, what is arguably an extension of the category of primary victim. Claimants who were not in physical danger or reasonable fear thereof have, in certain circumstances, been held to be primary, despite the decision in *White*. Thus a father who was wrongly informed that his baby had died and was given what he understood to be his dead baby to hold was classified as primary⁶² and so too was a person who claimed to have been falsely imprisoned because of his solicitor’s negligence.⁶³ It is not therefore only the involuntary participant who has presented the courts with difficult decisions regarding status. All these examples reinforce the conclusion that the law in this area has been too much driven by the Hillsborough decisions, largely determined by policy, and that the *Alcock* classification of victims is inappropriate in cases which do not result from large-scale tragedies. The “fear of floodgates” is not then an issue, and where the claimant is directly involved, although not in physical danger himself, he should not be treated as a secondary victim.

Dooley, which has never been expressly overruled, was determined on the basis of psychiatric harm being reasonably foreseeable and there is a strong argument for returning to this approach, at least in the case of an involuntary participant, and indeed other claimants directly involved in the incident, and not classifying them as either primary or secondary. This would surely mark a return to principle, the search for which was called off, according to Lord Hoffman in *White*.⁶⁴

Such an approach would, in practice, achieve the same result as that recommended by the Law Commission with regard to involuntary participants⁶⁵ and is reminiscent of Lord Bridge’s judgment in *McLoughlin v O’Brian*,⁶⁶ another case involving only one claimant. He opined that the defendant’s duty “must depend on reasonable foreseeability” and must be judged on a case-by-case basis rather than being pre-determined by a fixed category.⁶⁷ The Scottish Law Commission’s recommendations applicable, amongst others, to involuntary participants also refer to reasonable foreseeability of a medically recognised mental disorder⁶⁸ as being all that is required in such cases. Unless these draft bills become law it will be left to the judiciary to seek a way forward and it is submitted that cases like *W* and *Salter* represent a small step in the right direction.

⁶¹ *Galli-Atkinson v Seghal* [2003] EWCA Civ 697.

⁶² *Farrell v Avon Health Authority* [2001] Lloyd’s Rep Med 458.

⁶³ *McLoughlin v Jones* [2002] 2 WLR 1279.

⁶⁴ *Op cit.* n.1, at 511.

⁶⁵ *Op cit.* n.3, para.7.6.

⁶⁶ *Op cit.* n.47.

⁶⁷ *ibid.*, at 443.

⁶⁸ *Op cit.* n.55. The additional requirement that it be reasonably foreseeable that the act in question was likely to cause a person in the position of the victim to suffer harm of that kind is to clarify that there would be no requirement that the wrongdoer foresaw or was aware of the existence of the victim. This would be unlikely to be the position anyway in the case of an involuntary participant.

BOOK REVIEWS

Sean Coyle and George Pavlakos (eds.), Jurisprudence or Legal Science? A Debate about the Nature of Legal Theory. (Oxford, Hart Publishing, 2005, 174 pages (inc. index) ISBN 1-84113-504-6).

Jurisprudence or Legal Science? is a collection of contributions first presented at the “Workshop in Analytical Jurisprudence” held on 25-26 October 2002 at Queen’s University of Belfast. The specific question addressed by the papers contained in this volume is whether legal theory should be conceived of as an analysis aimed at clarifying the broad conceptual framework within which legal practice takes place (legal theory as legal science) or as a study engaging directly with the doctrinal debates which are of immediate concern for legal practitioners (legal theory as jurisprudence). A number of differences set the two enterprises apart. Legal science is a highly general and abstract form of conceptual analysis which aims at neutrality, austerity, detachment and objectivity; jurisprudence is an engaged project which assumes the unavailability of value-free explanations of law and so takes a position on the basic values that govern legal practices. Moreover, legal science is concerned with the nature of law in general; jurisprudence assumes that we cannot fruitfully reflect on the *general* characteristics of law but only on the *context-dependent* features of *given legal systems*.

If we focus on this last opposition (legal theory as a study of the nature of law versus legal theory as a situated enquiry of specific legal practices) we are in the best condition to both have a full grasp of the dilemma dealt with in *Jurisprudence or Legal Science?* and appreciate the originality of the contributions comprised in this volume. For, the essays in the volume can be roughly classified as either attempts to defend the ambitious thesis that a general philosophical enquiry into law’s nature is not only possible but also desirable or arguments for the impossibility of legal science.

The main assumptions underlying the more ambitious project are discussed by Jonathan Gorman and Carsten Heidemann. Gorman establishes a link between theoretical enquiry about law, on the one hand, and the concepts of truth, knowledge and understanding, on the other. On this view, we cannot grasp the structure and scope of legal theory unless we have a wider-ranging account of the notion of truth and knowledge. Heidemann, who discusses in detail Alexy’s version of the currently widespread view that legal systems raise a claim to correctness, connects the possibility of legal science to the cognitive status of legal norms. On his view, normativity, objectivity and mind-independence are all closely connected issues. If we follow Gorman and Heidemann, thus, we should conclude that when it is conceived of as a legal science, legal theory is open to the influences coming from the theory of truth, epistemology, ontology and theory of mind. As a result, legal theory is a truly philosophical enterprise.

This philosophical depth of legal theory is further explored and clarified by George Pavlakos, Veronica Rodriguez-Blanco and Robert Alexy. Pavlakos

argues for an idea of legal knowledge shaped by the thesis that normative properties are perceiver-dependant but nonetheless objective qualities of things. This approach, which is neither a kind of empiricism nor a form of essentialism, depicts legal knowledge as a rationalist normative activity, *i.e.* an interpretative process addressing entities provided with both an *a priori* component and an *a posteriori* component. Rodriguez-Blanco defends the possibility of legal science by means of a discussion of the metaphysical commitments of conceptual analysis. *Contra* Dworkin's criticism of Archimedeanism, Rodriguez-Blanco argues for the relevance of metaphysics in legal theory: far from consisting exclusively in debates on substantive questions the theory of law is profoundly concerned with metaphysical issues. Finally, Alexy states in the clearest possible way that the theoretical enquiry into law should be understood as a chapter of general philosophy. Philosophy is a "general and systematic reflection about what there is, what ought to be done or is good, and how knowledge about both is possible" (page 51). Legal theory raises the same questions although with respect to the more limited domain of law. Hence, it comprises four basic dimensions: a critical dimension, a normative dimension, an analytic dimension and a synthetic dimension. A careful exploration of all these dimensions will disclose the fundamental characteristics of law to us. As a result, a direct connection can be established between legal theory and nature of law: the very nature of law will progressively become apparent as we progress in our legal-theoretical studies.

For different reasons this conclusion is unwelcome by Philip Leith, John Morison and Sean Coyle. In their joint contribution Leith and Morison restate the importance of the anti-formalist critique of analytical legal theory, which is recognised to be both mistaken and exceedingly narrow in scope. For these reasons, analytical legal theory is constitutively unable to grasp the most important dimensions and transformations of the legal phenomenon such as the actual use of legal standards by practitioners and the passage from government to governance for example. This approach entails also a re-evaluation of empiricism: a scientific approach in legal theory requires us to undertake empirical studies into the practice of law. Accordingly, legal theory falls short of being a science as long as it aims at exploring the nature of law. To put it in a slogan, jurisprudence, not legal science, is the genuinely scientific approach to law. A preference for jurisprudence is also expressed by Coyle though on rather different theoretical grounds. On Coyle's view "the search for necessary or foundational truth in the context of law is itself misconceived" (page 22). There is nothing necessary in the concept of law since legal systems are historical and contingent products that reflect specific political and social conditions. Consequently even general theoretical approaches to law, such as Hart's for example, are better interpreted as attempts to elucidate the distinctive features of a legal system in a given historical period rather than as efforts to account for the universal and eternal properties of law. The argument for a deflationary approach to legal theory that Coyle deploys in his contribution is sophisticated and forceful. However, it ultimately leaves open the possibility for legal science. As long as comparative legal-theoretical studies are conceivable we cannot do without a comprehensive concept of law, namely an overarching idea that accounts for, but at the same time transcends, the basic characteristics of the specific historical experiences. For, in the absence of such an all-compassing category we would not be able to recognise the particular legal experiences

as different species of a same *genus*. They would rather appear to us as different kinds altogether between which no comparison can sensibly obtain. Comparative jurisprudence, then, requires a general account of law. In turn by hypothesis this general account can be provided by legal science only. As a consequence, far from being ruled out legal science is ultimately called for by jurisprudence itself.

Dr Stefano Berteà
University of Leicester

***A Star Chamber Court in Ireland: The Court of Castle Chamber, 1571-1641* by Dr. Jon G. Crawford (Four Courts Press, 2005. Illustrated hardback, 672 pages, £65.00)**

As the first monograph written on the subject of the Court of Castle Chamber *A Star Chamber Court in Ireland: The Court of Castle Chamber, 1571-1641* by Dr. Jon G. Crawford, who is the Director of International Education at Roanoke College in Salem, Virginia, represents a significant contribution to our knowledge of the institution, as well as to the individuals and events surrounding the Court. The work follows Dr. Crawford's extensive interest and unparalleled expertise in the area, chiefly embodied in "The Origins of The Court Of Castle Chamber; A Star Chamber Jurisdiction In Ireland", *American Journal Of Legal History*, 24 (1980) 22-55 and *Anglicising the Government of Ireland: the Irish Privy Council and the Expansion of Tudor Rule, 1556-1578* (1993), and is as scrupulous in its detail as it is complete in its coverage of the machinery of the court that was the counterpart of the English Star Chamber.

A Star Chamber Court in Ireland: The Court of Castle Chamber, 1571-1641 embraces the history of the Court, from its incipient prefiguration as the result of frustrated Henrician and Elizabethan reforms in the 1530s and again in the 1560s, to its proper establishment under Sir Henry Sydney in 1571, and then its shifting fortunes under consecutive Irish chief governors from Sydney's first successor Sir William Fitzwilliam to the final chief governor of the period Sir Thomas Wentworth. The character of the regimes of the other great viceroys like Sir Arthur Chichester (in particular, here, how the Court was utilised to prosecute cases of recusancy), as well as the associated administrations of other key figures such as Archbishop Adam Loftus and Attorney-General Sir John Davies are analysed in point, with their political context meticulously elucidated.

The author adroitly and elegantly demonstrates the emergence of the Court of Castle Chamber as an autonomous institution, and how this emergence came about against an environment of colonisation and rapid social change, and also one of busy if not entirely planned structural ordering. Equally Dr. Crawford shows, via the recorded examples of Castle Chamber cases, a drift away from a concentration on riot and violence focussed activities and towards a more miscellaneous and sophisticated jurisdiction. As the Monarch's peace was varyingly implemented, the regulation of a new range of legal concerns fell to the Court. From the litigants' perspective, the

plasticity of Castle Chamber's jurisdiction meant that a case which had little chance of being heard in the common law courts might more easily be brought before the Castle Chamber. There is clear evidence laid before the reader of the Court's systematic and ordered procedure. However, while Dr Crawford maintains the documented record of the Court confirms that it had become a tribunal that was keenly sought by litigants from all the regions of Ireland and from all religious and social groups, it is apparent that like other courts comparatively few of the community had genuine access to the Castle Chamber. It is plain that characteristics of prestige, wealth, masculinity and primogeniture were common to that element of the populace who can be shown to be associated, in some way, with the Court.

The final years of the Court, encompassing the period of the so-called Absolutism of Lord Deputy Wentworth (later the Earl of Strafford) from 1633 until 1638, and its end (including the impeachment and execution of Wentworth himself, and the breakdown of government in Ireland) from 1638 until 1641, are covered with apposite detail substantiating the opinion that in these years the Court of Castle Chamber was indeed utilised by the viceroy as an instrument of arbitrary power and political intimidation. By the end of this period, amid the general tumult, the Court receded and then vanished from operational activity – not so much with a bang as with nary even a whimper.

Across the water, in November of 1640, desperate and destitute Charles I had summoned Parliament and bent to its demands in exchange for the tender of supply. After the most immediate concerns of the Parliament's own life and fund raising methods had been addressed, the sights of the Commons turned upon the organs through which the King had applied his non-parliamentary government. The Commons established a committee to reform the Star Chamber, and two reform bills were considered, before it was reported to the House by Edmund Prideaux on the last day of May, 1641, that abolition was the sole option. Slight opposition to the measure was quickly quelled by Simonds D'Ewes. The bill for the abolition of Star Chamber, and most of the rest of the conciliar courts' jurisdictions, was passed and gained the assent of the defeated Charles on 5th July, 1641. The Star Chamber Abolition Act, 16 Car.I c.10, made no mention of the Court of Castle Chamber, nor indeed of any matter to do with the governance of Ireland. As noted by Dr. Crawford, the English Parliament did not interfere with Irish affairs until the crises of the 1640s. It is interesting to observe, all the same, that the absence of the Star Chamber by no means meant the end in England of political trials such as that of Prynne. Prior to its appropriation of power the Parliament had already established its own "star chamber" in the Committee for Examinations and Parliamentary Justice (in relation to which see W. Epstein, "The Committee For Examinations And Parliamentary Justice", *Journal Of Legal History*, 7,1 (1986) 3-22). Indeed, in 1641 the jurisdiction exercised by the Star Chamber was taken up by the court of King's Bench – with no real sense of recalcitrance.

It is important to call attention to a very important feature of the work, namely in the author's inclusion of extensive appendices. Because *A Star Chamber Court in Ireland* is the pioneering work on the topic of the Court, it is entirely appropriate for the author to have incorporated such wide-ranging source materials in this way. The appendices are four in number (although with internal sub-division within that number), and total some 168 pages

(mostly in a smaller than normal point size). The greatest part of the appendices, Appendix 1, is a transcription of British Library Additional MS 47, 172 being the entry book of orders and decrees of the Court of Castle Chamber from 1571 to 1621. The material represents the largest collection of the Court's functioning in its most raw form, and embodies the most substantial archival information available to scholars of the topic. Although the entry book of orders and decrees of the Court ends in 1621, Appendix 2 takes up the record, drawn now from Trinity College Dublin MS 852, for the years 1621-1632. Also from TCD MS 852, and reproduced in Appendix 3, is the Royal Commission of the Court of Castle Chamber of 1625. Appendix 4, extracted from various manuscript sources, details petitions and cases relating to the Court in its final years during the 1630s under the tenure of Thomas Wentworth as Lord Deputy. These appendices might indeed have formed a work on their own, so important is their substance, but here form an appropriate complement to the author's preceding work of historical interpretation and synthesis. The exertion of the Dr. Crawford in compiling and presenting the unprocessed materials of research in this way is an outstanding service to other scholars, and it is exciting to see the matter so exposed because there is still so much which can be wrought from these raw materials.

Other matters, relating to the form of *A Star Chamber Court in Ireland*, include a particularly thorough and extremely useful bibliography, which is suitably and conveniently divided into the major categories of sources. There is also, naturally, an index - which somewhat less helpfully seems to focus heavily upon the names of individual players and less heavily on institutional, procedural, thematic or conceptual issues. The volume, some 672 pages *in toto*, is published in a solid hardback by the Four Courts Press in association with The Irish Legal History Society. It is covered, handsomely, by a jacket incorporating a photograph of the seal of Lord Deputy Wentworth. This photograph is also reproduced in the fly leaves, and is sourced to Marsh's Library, Dublin, MS Z.3.2.6, item 109, p. 305. On the subject of illustration, the volume also contains sixteen plates, in black and white, each with commentary. The plates show many of the key actors in the Court's story - either in portraiture or in funereal effigy, some of the architecture and geography relevant to the Court, as well as an extract from each of BL Add MS 47, 172 and TCD MS 852.

Michael Stuckey
University of Glamorgan