

Northern Ireland Legal Quarterly

Volume 64 Number 2

EDITOR

PROFESSOR SALLY WHEELER



Queen's University
Belfast

School of Law

Contents

Public interest and the three dimensions of judicial review <i>Gordon Anthony</i>	125
Cartel criminalisation and due process: the challenge of imposing criminal sanctions alongside administrative sanctions within the EU <i>Dr Peter Whelan</i>	143
Carbon sequestration and the limits of regulation: a UK perspective <i>Mark Stallworthy</i>	165
The benefits of privacy-invading expression <i>Paul Wragg</i>	187
Women's work: human rights, gender and social class in Hungary at the turn of the twentieth century <i>István Pogány</i>	209
A more efficient derivative action system in China: challenges and opportunities through corporate governance theory <i>Dr Jingchen Zhao</i>	233
Beyond financialisation? How direct regulation will transform the future of financial market structures <i>Dieter Pesendorfer</i>	255

Public interest and the three dimensions of judicial review

GORDON ANTHONY*

Queen's University Belfast

Introduction

One of the most striking features of judicial review in the UK is the extent to which the term ‘the public interest’ is used by the courts when resolving different points of law. Historically, use of the term can perhaps be traced to the doctrine of public interest immunity that suffused the earlier conception of Crown privilege,¹ but it is since the procedural reforms of the late 1970s that the term has become almost a constant within the case law.² At one level, this has been a direct consequence of the reforms themselves, as aspects of the judicial review procedure have been said to safeguard the broader public interest in efficient public administration.³ But there has been a range of other factors that have equally led to an increased use of the term, whether in relation to procedural law or substantive law. These have included ongoing development of a common law principle of legality that emphasises the public interest in vindicating the rule of law in any case in which a public authority has acted unlawfully.⁴ Also relevant have been European standards on the limitation of fundamental rights, as well as international instruments like the Aarhus Convention that binds the UK and European Union (EU) on the shared matter of the environment.⁵

* Professor of Public Law, School of Law, Queen's University Belfast. This article is based upon the text of an inaugural lecture delivered at Queen's on 15 November 2012. My thanks are due to various colleagues and friends who commented upon earlier versions of the text: Jack Anderson, Pat Birkinshaw, Brice Dickson, Olivier Dubos, Carol Harlow, Tom Hickman, Peter Leyland, Kieran McEvoy, John Morison and Ting Xu. My thanks are also due to Leanne Cochrane who worked as a research assistant on an earlier Economic and Social Research Council-funded project that the article draws upon (The Public Interest in UK Courts: RES-062-23-1996). The responsibility for any errors is mine.

1 *Conway v Rimmer* [1968] AC 910; *Rogers v Secretary of State for the Home Department* [1973] AC 388; and *Al Rawi v Security Service* [2011] UKSC 34, [2012] 1 AC 531.

2 On the reforms see H W R Wade and C F Forsyth, *Administrative Law* (10th edn, OUP 2009), ch 18.

3 See Lord Woolf, *Protection of the Public: A New Challenge* (Steven & Sons 1990) 12–15; and, regarding delay, *O'Reilly v Mackman* [1983] 2 AC 237 and eg *Re Quinn's Application for Leave* [2010] NIQB 100 [14].

4 *R v Inland Revenue Commissioners, ex p National Federation of Self-Employed and Small Businesses Ltd* [1982] AC 617, 644, Lord Diplock. On the common law principle, see J Jowell, ‘The Rule of Law and its Underlying Values’ in J Jowell and D Oliver (eds), *The Changing Constitution* (7th edn, OUP 2011) ch 1.

5 The text of the Aarhus Convention is available at <www.unece.org/env/pp/documents/cep43e.pdf>.

Such increased references to the public interest reflect a constitutional reality that often places judicial review at the intersection of national, supranational and international legal norms. Although individual cases will inevitably be specific to their own facts and law, it is axiomatic that the wider body of judicial review case law now has three main dimensions – common law, European law, international law – and that these can coincide in particular cases. Where there is such coincidence, the courts may refer to the public interest as it is conceived of by the common law or by European law and/or international law. While such references can raise difficult questions about the nature of the public interest that is being appealed to – questions that become more complex when European law or international law is in issue – they provide a clear example of how far modern judicial review must accommodate different (and sometimes competing) legal norms. They also suggest that there is a need for nuance when considering the common law principle of legality that links the public interest to the rule of law, as the applicable legal norms will sometimes have been formulated at the European or international levels.

The corresponding aim of this article is to illustrate how references to the public interest can traverse the three dimensions of judicial review and to identify a conceptual framework for understanding those references. At the level of illustration, the article focuses on the example of public interest litigation whereby pressure groups, representative bodies and statutory organisations can enjoy liberal access to courts for purposes of bringing proceedings in their own names or intervening as third parties in ongoing disputes.⁶ Although the genesis of public interest litigation in the UK courts lies very much in the common law dimension of judicial review, it will be seen that European law and international law now also make demands about liberal access to the courts for a range of (not always connected or complementary) reasons. This is thus one area where there is a pronounced interaction of norms, and the article suggests that that interaction can be understood as an outworking of ‘global administrative law’.⁷ According to the global administrative law thesis, national courts are directly and indirectly involved in a number of legal networks that pursue accountability as a core value in contemporary global governance. While the thesis was originally concerned with the development of legal principles in an administrative space above the level of the state, it is now also associated with a global legality principle that is defined by a pluralist interaction of national, supranational and international legal orders.⁸ The article uses that thesis to examine instances of the intersection of norms in public interest litigation, where common law approaches to procedure have been key to facilitating the demands of European law and/or international law. The article likewise draws upon the idea of an emerging global legality principle to contextualise some UK court rulings on the substantive issues that have arisen in public interest cases, for instance, on the protection of fundamental rights and the limits to the discretion that may be exercised by public decision makers.

The analysis begins with a section that considers some of the different meanings of the public interest and the various ways in which the UK courts may refer to the term. It then considers the example of public interest litigation and, in that setting, the nature of the common law, European law and international law dimensions to judicial review. The final

6 *Axa General Insurance Ltd v Lord Advocate* [2011] UKSC 46, [2012] 1 AC 868; *Re E (A Child)* [2008] UKHL 66, [2009] 1 AC 536; and eg M Kirby, ‘Deconstructing the Law’s Hostility to Public Interest Litigation’ (2011) 127 LQR 537.

7 B Kingsbury, N Krisch and R Stewart, ‘The Emergence of Global Administrative Law’ (2005) 68 *Law and Contemporary Problems* 15.

8 J-B Auby, *La globalisation, le droit et l’État* (2nd edn, Montchrestien 2010).

section and the conclusion develop the point about the analytical worth of global administrative law and the interface between the public interest and the principle of legality.

Locating ‘the public interest’

Perhaps the first – self-evident – point that should be made about the public interest is that it is a concept that lacks precise definition and which has been the subject of polarised debate about its meaning and content.⁹ Although there have inevitably been a number of strands to the debate, the prevailing theme has been the legitimacy or otherwise of a presumed political community with shared values such as sustains the public interest. For instance, commentators who have been critical of the concept have noted the artificiality of ideas of political community and so on when suggesting that the concept has ‘fundamentally undemocratic’ implications for autonomous individuals.¹⁰ The argument here has been that the public interest is anti-libertarian in thrust as it can result in the otherwise valid preferences of individuals being subjected to those of falsely constructed majority communities.¹¹ However, such concerns about majoritarianism have been countered by alternative models that adopt a communitarian logic when accommodating individual interests within those of wider society.¹² While such modelling has begged obvious questions about how to gauge the public interest within polities that have a variety of cross-cutting individual and sectional interests, as well as those that may appear irreconcilable, it has nevertheless emphasised the existence of complementary interests among members of society. Answers to the obvious questions have therefore been found in ‘preponderance’, ‘unitary’ and ‘common interest’ theories that situate individual interests within that of a broader collective and emphasise the normative worth of the public interest.¹³

A further approach has centred upon the public interest as procedural in form. For Bozeman, writing about the interconnection between public values and the public interest, this approach marks the compromise position of those who could not ‘accept a normative view of public interest but [were] not ready to altogether abandon the concept’.¹⁴ Elemental to this approach is an understanding that the virtue of government lies not in its assumed ability to represent shared values but rather in ‘the multiplicity of points of access it affords for the manifold conflicting interests which necessarily arise’ in society.¹⁵ Government, on this approach, is therefore about reconciling interests through deliberation and debate and about having procedures that can yield agreed – or certainly acceptable – policies and outcomes. Indeed, while an emphasis on procedure also raises doubts about majoritarianism¹⁶ and about how to ensure coherent reasoning in respect of substantive choices,¹⁷ the procedural approach recognises that individuals necessarily interact with one another within the framework of broader society. On this view, the public

9 For overviews of the debates, see B Bozeman, *Public Values and Public Interest* (Georgetown UP 2007) ch 1; R C Box, ‘Redescribing the Public Interest’ (2007) 44 *Social Science Journal* 585; M Feintuck, *The ‘Public Interest’ in Regulation* (OUP 2004) ch 1; and V Held, *The Public Interest and Individual Interests* (Basic Books 1970).

10 The term was used by H R Smith: see *Democracy and the Public Interest* (University of Georgia Press 1960) 27.

11 A balance that led some to argue for the abolition of the term: see eg C E Cochran, ‘Political Science and “The Public Interest”’ (1974) 36 *Journal of Politics* 327.

12 See generally Held (n 9).

13 *Ibid.*

14 Bozeman (n 9) 93.

15 G A Schubert, *The Public Interest: A Critique of the Theory of a Political Concept* (Glencoe Free Press 1960) 136.

16 A Lijphart, *Thinking about Democracy: Power Sharing and Majority Rule in Theory and Practice* (Routledge 2007) especially chs 7 and 8.

17 See J Bell, ‘Public Interest: Policy or Principle’ in R Brownsword (ed), *Law and the Public Interest* (Franz Steiner Verlag 1993) 30.

interest is what emerges from deliberative processes that occur within democratically legitimated institutions.¹⁸

Whatever the strengths and weaknesses of such models, they permeate various well-known features of UK public law. This is true, for example, of the procedural approach and the UK constitution's emphasis on representative democracy.¹⁹ Although other (written) constitutions may note the nexus between 'the people' and 'the State'²⁰ – the latter being the 'formal expression and visible symbol of the "public interest"'²¹ – UK constitutional law has historically lacked such theorising and has centred, instead, on the place of the Crown and its relationship with the Westminster Parliament.²² This has led to some discussion about the nature of the interests that the Crown itself can represent,²³ but it has also, and more famously, resulted in the Westminster Parliament's elevation to a position of unique prominence within the constitution.²⁴ Moreover, while it is true that the doctrine of parliamentary sovereignty no longer enjoys unqualified judicial support²⁵ – causal considerations include devolution, Europeanisation and globalisation – the doctrine remains sourced in the idea that the Westminster Parliament is legally sovereign because its authority derives from the legitimating force of the electorate.²⁶ Seen in this way, the Westminster Parliament becomes a deliberative forum that brings together the competing interests within society; reconciles those interests through the legislative process; and consolidates the accepted outcomes in Acts that the courts regard as sovereign.

Appeals to ideas of 'community' can then be found in much of administrative law. At one level, this might be thought inevitable given the functions that administrative law is sometimes said to perform. Although the absence of a state tradition entails that UK law here too is significantly different from some other legal orders,²⁷ the body of administrative law that developed in the previous century was argued, by some, to centre upon the reinvention of society through the provision of public goods.²⁸ This, famously, is the realm of Harlow and Rawlings' 'green-light' theory of administrative law and it corresponds, at its height, with an understanding that public authorities exercise power on the basis of 'considerate altruism'.²⁹ Such altruism requires public bodies to consider the interests of individuals affected by their decisions, but it also, and just as importantly, means that 'public bodies have no rights or interests of their own and must exercise their powers . . . for the

18 See further I O'Flynn, 'Deliberating about the Public Interest' (2010) 16(3) *Res Publica* 299.

19 See generally P Leyland, *The Constitution of the United Kingdom: A Contextual Analysis* (2nd edn, Hart Publishing 2012) ch 5.

20 See eg Arts 4–11 of the Irish Constitution, 1937; and D Clarke, 'Nation, State and Nationality in the Irish Constitution' (1998) 16 *ILT* 252.

21 The term is W Friedmann's: see 'The Changing Content of Public Interest: Some Comments on Harold D Lasswell' in C J Friedrich (ed), *Nomos V: The Public Interest* (Atherton Press 1967) 80, 84.

22 N MacCormick, *Questioning Sovereignty: Law, State and Nation in the European Commonwealth* (OUP 1999) ch 3.

23 See eg M Loughlin, 'The State, the Crown and the Law' in M Sunkin and S Payne (eds), *The Nature of the Crown: A Legal and Political Analysis* (OUP 1999) 33; and T Nairn, *The Enchanted Glass: Britain and its Monarchy* (Routledge 1988).

24 J Goldworthy, *Parliamentary Sovereignty: Contemporary Debates* (OUP 2010); and *The Sovereignty of Parliament: History and Philosophy* (OUP 1999). Compare V Bogdanor, 'Imprisoned by a Doctrine: The Modern Defence of Parliamentary Sovereignty' (2012) 32 *OJLS* 179.

25 *R (Jackson) v Attorney-General* [2005] UKHL 56, [2006] 1 AC 262, 302 [102], Lord Steyn.

26 P Craig, *Public Law and Democracy in the United Kingdom and United States of America* (OUP 1990), ch 2.

27 J Allison, *A Continental Distinction in the Common Law: A Historical and Comparative Perspective on English Public Law* (OUP 2000).

28 C Harlow and R Rawlings, *Law and Administration* (3rd edn, CUP 2009), 31–37.

29 D Oliver, *Common Values and the Public–Private Divide* (Butterworths 1999) 5.

general good'.³⁰ The prevailing imagery is therefore of public bodies acting for the good of society as a whole and in the pursuit of only 'lawful and relevant grounds of public interest'.³¹ Indeed, it is for this reason that strict procedural requirements in judicial review have been said to be justified, as they are understood to provide a heightened degree of certainty in public administration that thereby benefits society as a whole.³² The courts have similarly referred to the 'interests of the community' when justifying, on grounds of 'public policy', a restrictive and remarkably durable line of case law on the negligence liability of public authorities.³³

'Red-light', or 'control', theories of administrative law also make mention of the public interest, although uses of the term here can vary.³⁴ At their core, such theories posit that the sovereign legislature has entrusted public authorities with power and that the primary function of administrative law is to ensure that the exercise of such power observes the limits set by Parliament.³⁵ This originally meant that control was synonymous with the *ultra vires* doctrine, but the courts have long since extended public law protections into the non-statutory field in disputes that have involved 'a matter of public interest in the sense that [they have] an impact on the public generally and not merely on an individual or group'.³⁶ This has been one aspect of the much commented upon expansion of the common law since the 1960s,³⁷ which has also involved increased recognition and protection of common law fundamental rights.³⁸ That protection initially took the form of 'anxious scrutiny' of executive decisions but it is now largely governed by the complementary principles that have emerged under the Human Rights Act 1998.³⁹ Prominent among these is the idea that there is a 'public interest' in recognising individual rights as this gives the law a normative grounding that can act as a brake on the abuse of governmental power.⁴⁰ However, also prominent is a line of reasoning that accepts that proportionate limitations may be placed upon qualified rights where the limitations pursue a legitimate ground of public interest within the meaning of the European Convention on Human Rights (ECHR) (for instance,

30 Oliver (n 29); and see further D Oliver, 'Psychological Constitutionalism' (2010) 69 CLJ 639, 640–44. But compare M Douglas, *How Institutions Think* (Syracuse UP 1986).

31 *R v Tower Hamlets LBC, ex p Chetnik Developments* [1988] AC 858, 872, Lord Bridge, quoting W Wade, *Administrative Law* (5th edn, OUP 1982) 355–56. See also M Elliott, 'Judicial Review's Scope, Foundations and Purposes: Joining the Dots' [2012] NZLR 75, 79.

32 Woolf (n 3) 12–15.

33 *Eg Brooks v Metropolitan Police Commissioner* [2005] UKHL 24, [2005] 1 WLR 1495, 1510, Lord Steyn. See, to similar effect, *Mitchell v Glasgow City Council* [2009] UKHL 11, [2009] 1 AC 874; *Jain v Trent SHA* [2009] UKHL 4, [2009] 1 AC 853; *Van Colle v Chief Constable of Hertfordshire; Smith v Chief Constable of Sussex* [2008] UKHL 50, [2009] 1 AC 225; *D v East Berkshire Community Health NHS Trust* [2005] UKHL 23, [2005] 2 AC 373; *Gorringe v Calderdale MBC* [2004] UKHL 15, [2004] 1 WLR 1057; and *Hill v Chief Constable of West Yorkshire* [1989] AC 53. But compare *Smith, Ellis and Allbutt v MOD* [2013] UKSC 41.

34 On red-light theory, see Harlow and Rawlings (n 28) 22–25.

35 See P Craig, *Administrative Law* (7th edn, Sweet & Maxwell 2012) 4–16.

36 *Re McBride's Application* [1999] NI 299, 310, Kerr J. See further G Anthony, *Judicial Review in Northern Ireland* (Hart Publishing 2008) 40–42.

37 See eg C F Forsyth (ed), *Judicial Review and the Constitution* (Hart Publishing 2000).

38 M Fordham, 'Common Law Rights' [2011] 16 JR 14.

39 *Bugdaycay v Secretary of State for the Home Department* [1987] AC 514, 531, Lord Bridge; T Hickman, *Public Law after the Human Rights Act* (Hart Publishing 2010); and B Dickson, *Human Rights and the UK Supreme Court* (OUP 2013) ch 2.

40 See eg *Derbyshire CC v Times Newspaper Ltd* [1993] AC 534 discussing the public interest in freedom of expression. On rights and the abuse of power see Craig (n 35) 16–27.

national security, public order, health or morals etc).⁴¹ Such reasoning, which has some parallel under the derogation provisions of the Treaty on the Functioning of the European Union (TFEU)⁴² and in the EU Charter of Fundamental Rights,⁴³ again appeals to the imagery of community by noting that ‘the fundamental rights of individuals are of supreme importance but . . . are not unlimited: we live in communities of individuals who also have rights’.⁴⁴ There is, in the result, a dual use of the public interest in rights discourse that carries a presumption in favour of rights but which acknowledges ‘the duties of everyone to the community and . . . [the need] . . . to secure and protect respect for the rights of others’.⁴⁵

Attempts to balance the rights of individuals with those of the community can, however, lead to criticisms of majoritarianism, for instance, where the balance is too often struck in favour of the community or, more pointedly, legislative or executive choices.⁴⁶ Here, the common law has previously sought to protect individuals by noting that their rights can intersect with the interests of third parties and that those third-party interests, in turn, may form part of the broader public interest.⁴⁷ Such complementarity of interests suggests a high degree of nuance in judicial reasoning, but it must nevertheless cede, in a concrete dispute, to a choice between the rights of the individual and the interests of the wider public. Faced with that choice, it is well known that courts will regard the context to a dispute as ‘everything’ and that the balance to be struck will depend on the rights that are involved and the type of decision or act that affects them.⁴⁸ To borrow from Laws LJ’s much-cited judgment in *International Transport Roth*, the courts can thus be expected to be more interventionist where individual rights are stated in terms that are unqualified and where the decision under challenge has been taken by an unelected official or even a minister who is also answerable to a legislature.⁴⁹ However, where the choice is that of the legislature itself - whether at Westminster or at the devolved levels - judicial restraint may, though not necessarily will, be writ large.⁵⁰

Of course, the scope for challenges to legislative choices, or certainly those of the Westminster Parliament, is one of the most prominent out-workings of the European law

41 See, illustratively, Arts 8-11 ECHR. But see A McHarg, ‘Reconciling Human Rights and the Public Interest: Conceptual Problems and Doctrinal Uncertainty in the Jurisprudence of the European Court of Human Rights’ (1999) 62 MLR 671.

42 Eg Art 36 TFEU: ‘The provisions of Articles 34 and 35 shall not preclude prohibitions or restrictions on imports, exports or goods in transit justified on grounds of public morality, public policy or public security; the protection of health and life of humans, animals or plants; the protection of national treasures possessing artistic, historic or archaeological value; or the protection of industrial and commercial property. Such prohibitions or restrictions shall not, however, constitute a means of arbitrary discrimination or a disguised restriction on trade between Member States.’

43 Viz. in Art 52. On the Charter’s status and reach in UK law see Case C-411/10, *NS v Home Secretary* [2012] 2 CMLR 9 and eg *Rugby Football Union v Viagogo Ltd* [2012] UKSC 55; [2012] 1 WLR 3333.

44 *Brown v Stott* [2003] 1 AC 681, 707–08, Lord Steyn.

45 *Ibid.*

46 See *Re P* [2008] UKHL 38, [2009] 1 AC 173; and eg L Lustgarten and I Leigh, *In from the Cold: National Security and Parliamentary Democracy* (Clarendon Press 1994). For a critical analysis of early case law under the Human Rights Act 1998, see R A Edwards, ‘Judicial Deference Under the Human Rights Act’ (2002) 65 MLR 859.

47 *Singh v Immigration Appeal Tribunal* [1986] 1 WLR 910.

48 *R v Home Secretary, ex p Daly* [2001] UKHL 25; [2001] 2 AC 532, 548, Lord Steyn

49 *International Transport Roth GmbH v Home Secretary* [2003] QB 728, 765ff.

50 *Ibid.*; and see *James v UK* (1986) 8 EHRR 123, 142, ECtHR recognising in the context of legislation interfering with property rights that ‘the notion of “public interest” is necessarily extensive’. See also *Axa General Insurance Ltd v Lord Advocate* [2011] UKSC 46, [2012] 1 AC 868. But, for an instance of intervention, see *R (F) v Secretary of State for the Home Department* [2010] UKSC 17, [2011] 1 AC 331 (Sexual Offences Act 2003, s 82, declared incompatible with Art 8 ECHR).

dimension of judicial review.⁵¹ The significance of that dimension is examined in more detail below, but the point to be noted at this stage is that challenges to the public interest choices of the Westminster Parliament, at least within the procedural meaning of the term, are underpinned by norms that have been defined in other deliberative and judicial fora. While this begs the question of the model of the public interest that predominates in those fora – there have historically been concerns about the quality of deliberation in the EU institutions⁵² – the intersection of norms nevertheless means that the choices of the Westminster Parliament no longer automatically prevail (at least within the context of EU membership).⁵³ That reality has since led to the European Union Act 2011 that refines the basis of EU membership, but, even though the Act modifies the mechanisms for further transfers of power to the EU, it does nothing to limit the current reach of EU law or, indeed, the common law's ascription of primacy to it.⁵⁴ More tellingly, some commentators have doubted whether it is even feasible to ring-fence national powers in the manner envisaged by the European Union Act 2011, precisely because contemporary patterns in governance mean that there are unavoidable points of intersection between national, supranational and international norms.⁵⁵ There is, in the result, a 'divided sovereignty' in the UK whereby claims to the public interest can be made both from within and outwith the domestic constitution.⁵⁶

Public interest litigation and the three dimensions of judicial review

Many of the above uses of the public interest have been displayed in the context of public interest litigation in UK courts. Public interest litigation, for these purposes, has been described above in terms of liberal access to courts for pressure groups, representative bodies and statutory organisations that wish either to bring proceedings in their own names or to intervene as third parties in ongoing disputes.⁵⁷ Within this, pressure groups etc. and the courts may have different reasons for favouring public interest litigation, even if their reasons may overlap on the facts of a given case. For instance, pressure groups and so on may wish to bring, or intervene in, proceedings precisely because they claim to be acting on a matter of public interest such as the environment, poverty, or fundamental rights.⁵⁸ However, while this links public interest litigation to wider policy areas and governmental

51 *R v Secretary of State for Transport, ex p Factortame Ltd (No 2)* [1991] 1 AC 603. See, in respect of the legislative choices of the devolved legislatures, Government of Wales Act 2006, ss 94 and 108; Northern Ireland Act 1998, s 6; and Scotland Act 1998, s 29.

52 See generally P Craig and G de Búrca, *EU Law: Text Cases and Materials* (5th edn, OUP 2011) chs 5 and 6. And, for an earlier account of the EU's processes, see A Moravcsik and A Sangiovanni, 'On Democracy and "Public Interest" in the European Union' in W Streeck and R Mainz (eds), *Die Reformierbarkeit der Demokratie. Innovationen und Blockaden* (Campus Verlag 2002) 122.

53 Compare Human Rights Act 1998, ss 3–4, and eg Representation of the People Act 1983, ss 3–4; *Hurst v UK (No 2)* (2006) 42 EHRR 41; and *Smith v Scott* (2007) CSH 9.

54 *Thoburn v Sunderland CC* [2003] QB 151. For commentary on the Act, see P Craig, 'The European Union Act 2011: Locks, Limits, and Legality' (2011) 48 CMLR 1915 and M Gordon and M Dougan, 'The United Kingdom's European Union Act 2011: "who won the bloody war anyway?"' (2012) 37 ELRev 1.

55 The leading commentary on the point remains M Hunt, *Using Human Rights Law in English Courts* (Hart Publishing 1997) chs 1–3; and for the patterns in governance see A Hurrell, *On Global Order: Power, Values, and the Constitution of International Society* (OUP 2007).

56 For the term 'divided sovereignty' see *R (Jackson) v Attorney-General* [2006] 1 AC 262, 302 [102], Lord Steyn.

57 See further Kirby (n 6). For discussion of what can constitute a 'public interest group', see JUSTICE/Public Law Project, *A Matter of Public Interest: Reforming the Law and Practice on Interventions in Public Interest Cases* (JUSTICE/Public Law Project 1996) 8–9.

58 See J Miles, 'Standing in a Multi-Layered Constitution' in N Bamforth and P Leyland (eds), *Public Law in a Multi-Layered Constitution* (Hart Publishing 2003) ch 15.

choices that affect society or ‘the community’, the courts have emphasised that they facilitate proceedings brought by pressure groups and so on because of the need to ensure that governmental decision makers act lawfully. Although this can give rise to controversy when judicial decision-making fringes upon matters of policy,⁵⁹ the courts have noted that there can be merit in constraining decision makers even in those cases where no individual has been directly affected by a decision.⁶⁰ On this (red-light) view, the public interest lies in the rule of law and a robust common law principle of legality.

In terms of the three dimensions of judicial review, the move towards public interest litigation has historically had its origins very much in the common law dimension. However, as the European law and international law dimensions of judicial review have become ever more prominent, the dynamics that drive public interest litigation have changed. This has been true not just in terms of the matters of the public interest that pressure groups and so on may appeal to – fundamental rights and the environment again providing obvious examples – but also in terms of related demands about the need for pressure groups and so on to have liberal access to the courts. It is thus here that the intersection of norms is to be found, and where global administrative law enjoys an explanatory force.

THE COMMON LAW DIMENSION

The common law’s approach to public interest litigation centres upon procedural matters of standing, costs and third-party interventions. In relation to standing, it is well known that the courts have long interpreted the ‘sufficient interest’ requirement in judicial review broadly and that groups such as the Child Poverty Action Group, Greenpeace and the World Development Movement have enjoyed liberal access to the courts⁶¹ (some organisations also have a statutory power to bring proceedings).⁶² This is where debate about the role of the courts has been at its most pronounced, as a liberal standing regime has opened the courts to applications that may ultimately be motivated by broader policy concerns rather than particular points of law. While few would doubt that there is a moral force to the causes associated with many pressure groups and so on, the doctrinal argument is that such groups are partisan and that there will often be rational, competing views about how to address the claimed matter of public interest. For some commentators, the inevitable concern is that the courts can thereby become a forum for disputes that a procedural model of the public interest would place elsewhere.⁶³

Costs in the context of public interest litigation are associated with so-called protective costs orders (PCOs) that cap the liabilities of applicants in proceedings which may

59 See generally C Harlow, ‘Public Law and Popular Justice’ (2002) 65 MLR 1.

60 *R v Somerset CC, ex p Dixon* [1998] EnvLR 111, 121, Sedley J; and *R v Secretary of State for Foreign and Commonwealth Affairs, ex p World Development Movement Ltd* [1995] 1 WLR 386, 395, Rose LJ.

61 See, respectively, *R (Child Poverty Action Group) v Secretary of State for Work and Pensions* [2012] EWHC 2579 (Admin), [2012] ACD 109; *R v HM Inspectorate of Pollution, ex p Greenpeace Ltd (No 2)* [1994] 4 All ER 329; and *R v Secretary of State for Foreign and Commonwealth Affairs, ex p World Development Movement Ltd* [1995] 1 WLR 386. Compare, most famously, *R v Secretary of State for the Environment, ex p Rose Theatre Trust Co* [1990] 1 QB 504. For commentary, see P Leyland and G Anthony, *Textbook on Administrative Law* (7th edn, OUP 2013) 201–06; and R Rawlings, ‘Modelling Judicial Review’ (2008) 61 CLP 95, 98–103.

62 See eg Northern Ireland Act 1998, s 71(2A)–(2C), and *Re Northern Ireland Human Rights Commission’s Application* [2012] NIQB 77. See too the powers of the Equality and Human Rights Commission in the Equality Act 2006, s 30, and eg *R (Equality and Human Rights Commission) v Prime Minister* [2011] EWHC 2401; [2012] 1 WLR 1389.

63 Harlow (n 59); and J A G Griffiths, ‘The Political Constitution’ (1979) 42 MLR 1.

otherwise be discontinued for financial reasons.⁶⁴ As will be seen below, this is one area in which there is a pronounced interplay between the three dimensions of judicial review, specifically in the context of environmental proceedings. However, the common law itself does not limit PCOs to the environmental setting, and they may be granted in any proceedings that raise an issue of ‘general public importance’ which ‘the public interest’ requires the court to hear (proceedings for these purposes may be brought either by a representative body and so on, or by an individual).⁶⁵ While these are not the sole criteria that the courts consider on an application for a PCO – others include the respective financial positions of the parties; whether the applicant has any private interest in the proceedings; and whether the applicant’s lawyers are acting on a pro bono basis⁶⁶ – they are an obvious correlate of the law that has developed around standing. Indeed, to the extent that it was initially thought that PCOs would issue only ‘exceptionally’, the Court of Appeal in England and Wales has since said that this is not an additional criterion and that the judicial task in any case remains one of isolating issues of ‘general public importance’ that are in the public interest.⁶⁷ Whether that can be done is ‘a question of degree and a question which the ... [case law] ... would expect judges to be able to resolve’.⁶⁸

Third-party, or public interest, interventions obviously engage pressure groups and so on in a different, less costly exercise, as an intervener purports either to supplement the arguments advanced by one of the parties to the proceedings and/or to bring wider points of law to the attention of the court.⁶⁹ In this context, the partisan nature of the intervener should be less remarkable than it (potentially) is in respect of standing, as the legal issues in the case will have been identified by a private individual who has been affected by the (in)action of a public authority (unless, that is, the proceedings have been initiated by another representative body).⁷⁰ Whether interventions will be accepted is ultimately a matter for the discretion of the court⁷¹ – albeit that some groups enjoy a statutory power to apply to intervene⁷² – and the judicial understanding is that an intervener’s ‘fund of knowledge or particular point of view will enable [it] to provide the [court] with a more rounded picture than it would otherwise obtain’.⁷³ This has been said to require that

64 The leading case remains *R (Corner House Research) v Secretary of State for Trade and Industry* [2005] EWCA Civ 192, [2005] 1 WLR 2600, as revisited in *R (Compton) v Wiltshire Primary Care Trust* [2008] EWCA Civ 749, [2009] 1 WLR 1436. And see Liberty, *Litigating the Public Interest: Report of the Working Group on Facilitating Public Interest Litigation* (Liberty 2006).

65 *R (Corner House Research) v Secretary of State for Trade and Industry* [2005] EWCA Civ 192, [2005] 1 WLR 2600, 2624 [74].

66 *Ibid.* Although, on the issue of private interest, see *Re Thompson’s Application* [2010] NIQB 38; and, on pro bono representation in England and Wales, see the Legal Services Act 2007, s 194, and the Legal Aid, Sentencing, and Punishment of Offenders Act 2012, s 61.

67 *R (Compton) v Wiltshire Primary Care Trust* [2008] EWCA Civ 749, [2009] 1 WLR 1436, 1446 [24], Waller LJ.

68 *Ibid.*; and for application, see eg *R (Technoprint plc) v Leeds City Council* [2009] EWHC 3832 (Admin). But compare *R (Action Against Medical Accidents) v General Medical Council* [2009] EWHC 2522 (Admin).

69 *Re Northern Ireland Human Rights Commission* [2002] UKHL 25, [2002] NI 236 noting, among other things, the different roles of an intervener and *amicus curiae*. On interveners, see further M Fordham, ‘Public Interest Interventions in the Supreme Court: Ten Virtues’ [2010] 15 JR 18; and JUSTICE/Public Law Project (n 57).

70 As in eg *Family Planning Association of Northern Ireland v Minister for Health, Social Services and Public Safety* [2005] NI 188, interventions made by, among others, the Roman Catholic Church and the Society for the Protection of the Unborn Child.

71 See eg RSC r 26.

72 Eg *Re Northern Ireland Human Rights Commission* [2002] NI 236 and Northern Ireland Act 1998, s 71(2B). See too the Commissioner for Children and Young People (NI) Order 2003, art 14(1)(b).

73 *Re E (A Child)* [2008] UKHL 66, [2009] 1 AC 536, 542 [2], Lord Hoffmann. On the limitations of traditional *inter partes* proceedings, see J Allison, ‘The Procedural Reason for Judicial Restraint’ [1994] PL 452.

interveners avoid mere repetition of points which have already been made by the parties and that they focus, instead, on additional matters that will assist the court.⁷⁴

THE EUROPEAN LAW DIMENSION

One area of law in which interventions have featured particularly prominently – and which marks a crossover between the common law and European law dimensions of judicial review – is that concerned with human rights.⁷⁵ Although human rights proceedings can also include arguments about fundamental rights at common law,⁷⁶ the majority of cases are now heard under the Human Rights Act 1998 as read with the relevant provisions of the ECHR.⁷⁷ In that latter setting, public interest litigation has assumed a more truncated meaning, as section 7 of the Human Rights Act 1998 precludes actions by representative bodies and so on save where they themselves are ‘victims’ within the meaning of Article 34 ECHR or where statute permits such actions.⁷⁸ That said, the fact that pressure groups etc. are able to intervene as third parties allows them to make submissions about the interface between national legislative and administrative choices and the body of case law that has been developed by the European Court of Human Rights (ECtHR). While that case law famously includes a ‘margin of appreciation’ doctrine that gives national authorities considerable latitude in their treatment of qualified rights, interveners can still address the normative importance of rights and whether the ECHR recognises a public interest justification for limiting rights in a particular case.⁷⁹ In some instances, interventions may influence a court towards the conclusion that even an Act of the Westminster Parliament cannot be reconciled with particular rights and that a ‘declaration of incompatibility’ should issue.⁸⁰

Litigation within the rubric of EU law is more complex, largely because of the role that the national courts play within the EU’s broader judicial architecture. Here, UK courts perform an essentially dual function whereby they can hear public interest challenges to acts of the EU institutions as well as to national legislative and administrative choices that are argued to contravene EU law. In terms of challenges to acts of the EU institutions, it is well established that pressure groups etc. have only restricted rights of access to the General Court and that proceedings should instead be brought in national courts, which may subsequently refer a matter of legality to the Court of Justice of the European Union (CJEU) under Article 267 TFEU.⁸¹ While this is a state of affairs that has been much criticised⁸² – the scope for direct actions has since also been partly liberalised by the Treaty

74 *Re E (A Child)* [2008] UKHL 66, [2009] 1 AC 536, 542 [2], Lord Hoffmann. See too *Re White’s Application* [2000] NI 432, 445.

75 See generally E Metcalfe, *To Assist the Court: Third Party Interventions in the UK* (JUSTICE 2009).

76 See eg *Ahmed v HM Treasury* [2010] UKSC 2 and 5, [2010] 2 AC 534, considered below.

77 See both Hickman and Dickson (n 39).

78 See eg *Re Committee for the Administration of Justice’s Application* [2005] NIQB 25. For an example of legislation giving bodies the power to bring proceedings see Northern Ireland Act 1998, s 71(2A)–(2C), and *Re Northern Ireland Human Rights Commission’s Application* [2012] NIQB 77.

79 But, for a critical account of the ECtHR case law, see McFarg (n 41).

80 Eg *A v Secretary of State for the Home Department* [2004] UKHL 56, [2005] 2 AC 68, declaration of incompatibility between the Anti-terrorism, Crime and Security Act 2001, ss 23–24, and Arts 5 and 14 ECHR, Liberty intervening.

81 T de la Mare and C Donnelly, ‘Preliminary Rulings and EU Legal Integration: Evolution and Stasis’ in P Craig and G de Búrca (eds), *The Evolution of EU Law* (2nd edn, OUP 2011) 363, 385–86. Direct actions are now governed by Art 263(4) TFEU.

82 See, most tellingly, Advocate General Jacob’s Opinion in Case C-50/00 *Unión de Pequeños Agricultores v Council* [2002] ECR I-6677.

of Lisbon⁸³ – it corresponds with a constitutional arrangement whereby national courts augment the work of the CJEU and the principles it has developed.⁸⁴ That arrangement also informs the role that UK courts play when hearing challenges to national legislative and administrative choices, where the supremacy and direct effect doctrines require that EU law prevails over any conflicting norm of national law.⁸⁵ The resolution of this latter type of challenge may again require a national court reference to the CJEU under Article 267 TFEU,⁸⁶ and it may at other times require that the national court engages in *de facto* review of the constitutionality of Acts of the Westminster Parliament.⁸⁷ When engaged in review of that kind, the national court will essentially ask whether a deliberative choice of the Westminster Parliament is consistent with externally defined constitutional norms.⁸⁸

The outstanding example of such review in public interest proceedings remains *R v Secretary of State for Employment, ex parte Equal Opportunities Commission*⁸⁹ (EOC). The EOC here argued that various provisions of the Employment Protection (Consolidation) Act 1978 were contrary to some of EU law's non-discrimination rules because they granted preferential employment protection rights to full-time workers, a majority of whom were men, as opposed to part-time workers, a majority of whom were women. Granting the application for judicial review, the House of Lords held: that the EOC had had 'sufficient interest' to bring the proceedings because of its statutory duty to work towards the elimination of discrimination; that the seminal *Factortame* case had made possible the review of Acts of the Westminster Parliament in the EU context; and that the Secretary of State had been unable to offer any objective justification for the discriminatory measures under challenge. On this outcome, the common law's open approach to standing elided with the core doctrines of EU law to produce a declaration that the Act of 1978 was incompatible with EU law's sex equality rules.

THE INTERNATIONAL LAW DIMENSION

The nature of the international law dimension can be illustrated with reference to the Aarhus Convention that was noted, above, in the introduction. This is a Convention that is of application only within the broader European region (including at the level of the EU) but which was drafted under the auspices of the United Nations Economic Commission for Europe (UNECE), drew inspiration from principle 10 of the United Nations' Rio Declaration on Environment and Development, and has been described as of 'global' significance.⁹⁰ Its central premise is the shared public interest in adequate protection of the environment, and it contains wide-ranging provisions on access to information on environmental matters, public participation in decision-making, and access to justice where

83 Art 263(4) TFEU. But, for continued criticism, see S Marsden, 'Direct Public Access to EU Courts: Upholding Public International Law via the Aarhus Convention Compliance Committee' (2012) 81 *Nordic Journal of International Law* 175.

84 *de la Mare and Donnelly* (81).

85 Case 16/62 *Van Gend en Loos v Nederlandse Administratie der Belastingen* [1963] ECR 1; Case 6/64 *Costa v ENEL* [1964] ECR 585; and Case 106/77 *Amministrazione delle Finanze dello Stato v Simmenthal SpA* [1978] ECR 629.

86 As in eg Case C-388/07 R (*Incorporated Trustees of the National Council on Ageing (Age Concern England) v Secretary of State for Business, Enterprise and Regulatory Reform*) [2009] ECR I-1569.

87 *R v Secretary of State for Transport, ex p Factortame Ltd (No 2)* [1991] 1 AC 603.

88 *R v Secretary of State for Transport, ex p Factortame Ltd (No 2)* [1991] 1 AC 603 and *Thoburn v Sunderland CC* [2003] QB 151.

89 [1995] 1 AC 1.

90 R McCracken and G Jones, 'The Aarhus Convention' [2003] JPL 802, quoting the former UN Secretary General, Kofi Anan.

public authority decisions are under challenge.⁹¹ Within this, the role that pressure groups etc. can play is explicitly recognised, albeit as subject to each party's procedural rules on standing,⁹² and the Aarhus Convention requires that remedies should be 'adequate and effective' and access to them 'fair, equitable, timely and not prohibitively expensive'.⁹³ Questions of adherence are overseen by a Compliance Committee that can act on a submission that is made by a party to the Aarhus Convention, a referral that is received from the Secretariat to UNECE, or a communication from a member of the public (including a pressure group).⁹⁴

Implementation of the Aarhus Convention in the UK has been achieved primarily through the modification of EU Directives and corresponding national implementing measures.⁹⁵ In terms of its content, the 'prohibitively expensive' provision has given rise to the most case law both in the UK courts themselves and, as regards the UK as a party to the Aarhus Convention, before the Compliance Committee.⁹⁶ Much of that case law has inevitably centred upon the domestic approach to PCOs and the Supreme Court recently referred to the CJEU the question of how 'prohibitive expense' is to be assessed. The issue arose in *Edwards*,⁹⁷ which was a challenge to a grant of planning permission brought by a third-party private individual who had applied unsuccessfully for a PCO in proceedings in the House of Lords and who then challenged a costs order against her when the appeal was dismissed (the House of Lords had in the meantime been replaced by the Supreme Court). In considering how compliance with the Aarhus Convention was to be observed, the Supreme Court noted the competing options presented by a wholly *subjective* approach to costs (what can the applicant be expected to pay given his/her/its financial position?) as opposed to a wholly *objective* approach (what might the 'ordinary' member of the public be expected to pay in the circumstances of the case?). While Lord Hope thought that the balance might 'lie in favour of the objective approach',⁹⁸ he was ultimately of the view that the law was uncertain and that the meaning of the relevant Directives should be referred to the CJEU under Article 267 TFEU. The CJEU, in its ruling of 11 April 2013, stated that the appropriate test should include elements of both approaches.⁹⁹

91 For a wider, contextual account, see G Smith, *Deliberative Democracy and the Environment* (Routledge 2003).

92 Art 2(5).

93 Art 9(4). See further J Maurici, 'Aarhus and Access to Justice' [2011] 16 JR 253.

94 See <www.unece.org/env/pp/ccbackground.html>; and M Macchia, 'Legality: The Aarhus Convention and the Compliance Committee' in S Cassese et al (eds), *Global Administrative Law: The Casebook* (3rd edn, Institute for International Law and Justice, NYU 2012) III.A.1.

95 See eg Dir 2003/35/EC on public participation, implementing the access to justice requirements by inserting Art 10a into Dir 85/337/EEC on Environmental Impact Assessments and art 15a into Dir 96/61/EC on Integrated Pollution Prevention and Control. For corresponding modification of domestic regulations, see eg the Town and Country Planning (Environmental Impact Assessment) (Amendment) Regulations 2006, SI 2006/3295. And on the Convention's requirements on access to information see Council Dir 2003/4/EC and the Environmental Information Regulations 2004, SI 2004/3391 and the Environmental Information (Scotland) Regulations 2004, SSI 2004/520.

96 See Maurici (n 93) 253–74.

97 R (*Edwards*) v *Environment Agency* (No 2) [2010] UKSC 57, [2011] 1 WLR 79.

98 [2010] UKSC 57, [2011] 1 WLR 79, 91, para 31.

99 Case C-260/11 R (*Edwards*) v *Environment Agency*, NYR, available at <www.bailii.org/eu/cases/EUECJ/2013/C26011.html>. And, on costs in England and Wales, see now also the Civil Procedure (Amendment) Rules 2013, SI 2013/262.

Public interest litigation, legality and global administrative law

Edwards provides a clear example of how the three dimensions of judicial review can coincide in public interest litigation, as the Aarhus Convention had prompted the modification of the EU Directives, which then raised issues about the nature of PCOs in national law. At its most obvious, that coincidence took form in the range of legal requirements on the protection of the environment that the applicant relied upon when bringing the proceedings on that matter of public interest. However, the coincidence also had a more complex quality that relates to the common law's principle of legality and the courts' emphasis on the public interest in vindicating the rule of law. That point of emphasis had already led to a liberal procedural regime in UK courts independent of the demands of European law and international law, and the Aarhus Convention, in that sense, did not entail any reinvention of national principle and practice. Nevertheless, the fact that *Edwards* involved the courts in enforcing rules of law that had been formulated outside the UK constitutional order also revealed how matters of legality can be determined at the intersection of national, supranational and international norms. So, what is the nature of the contemporary legality principle that may be involved in such public interest litigation? And what is the role of the common law principle of legality in cases that transcend the three dimensions of judicial review?

The approach that literature on global administrative law takes to such questions starts from the premise that public power is exercised at a number of cross-cutting levels in contemporary society – mainly national, supranational and international – and that corresponding legal constraints emerge from within the different legal orders at those levels.¹⁰⁰ Within that, the global administrative law thesis regards relations between the legal orders as heterarchical in form, and it borrows from 'constitutional pluralism' when noting a degree of equivalence between the values that define the various orders.¹⁰¹ This clearly departs from any account that would attribute a primary (hierarchical) role to international law, and it posits a number of interesting consequences that flow from the interaction of norms. One is that global administrative law becomes inextricably linked to constitutional values that exist in different legal orders and which include, most prominently, fundamental rights standards.¹⁰² A second, admittedly exceptional, consequence is that there is scope for one legal order to reject a norm that has originated in another order, for the reason that the norm offends a constitutional value within the receiving system. This is famously – and controversially – what happened in the *Kadi* case when the CJEU reviewed the legality of

100 There is a wide body of literature that touches upon this point and others. See, most prominently, Cassese et al (n 94); G Anthony, J-B Auby, J Morison and T Zwart (eds), *Values in Global Administrative Law* (Hart Publishing 2011); J-B Auby (n 8); D Dyzenhau, 'The Concept of (Global) Administrative Law' [2009] *Acta Juridica* 3; B Kingsbury, 'The Concept of "Law" in Global Administrative Law' (2009) 20 *EJIL* 23; Kingsbury et al (n 7); S Cassese, 'Administrative Law without the State? The Challenge of Global Regulation' (2005) 37 *New York University Journal of International Law and Politics* 663; G della Cananea, 'Beyond the State: The Europeanization and Globalization of Procedural Administrative Law' (2003) 9 *EPL* 565; A Aman Jr, 'The Limits of Globalization and the Future of Administrative Law: From Government to Governance' (2001) 8 *Indiana Journal of Global Legal Studies* 379; and C Harlow, 'European Administrative Law and the Global Challenge' in P Craig and G de Búrca (eds), *The Evolution of EU Law* (OUP 1999) 261. Compare A von Bogdandy, P Dann and M Goldmann, 'Developing the Publicness of Public International Law: Towards a Legal Framework for Global Governance Activities' (2008) 9 *German Law Journal* 1375.

101 The leading account remains N Walker, 'The Idea of Constitutional Pluralism' (2002) 65 *MLR* 317. See too C Harlow, 'Three Phases in the Evolution of EU Administrative Law' in Craig and de Búrca (n 81) 439; and M P Maduro, 'Contrapunctual Law: Europe's Constitutional Pluralism in Action' in N Walker (eds), *Sovereignty in Transition* (Hart Publishing 2003) 501. Compare N Krisch, *Beyond Constitutionalism: The Pluralist Structure of Post-National Law* (OUP 2010).

102 G Anderson, *Constitutional Rights after Globalisation* (Hart Publishing 2005).

EU Regulations that implemented UN Resolutions that had been adopted without regard for, among other things, fair hearing guarantees (the Resolutions provided for the freezing of the financial assets of terror suspects whose names had been added to a sanctions list by a UN Committee).¹⁰³ While the CJEU was not competent to rule directly on the legality of the UN Resolutions, it emphasised its role as the guardian of values within the EU's constitutional space and on that basis held that the EU Regulations were invalid.¹⁰⁴ This is thus a third consequence of constitutional pluralism: the emergence of a legality principle that can include an indirect judicial control of decisions that have been taken beyond the jurisdiction of the court giving judgment.

Of course, it is the nature and content of that legality principle that is of most interest here. Normatively, the principle is grounded in global administrative law's pursuit of the value of accountability in global decision-making,¹⁰⁵ where it emphasises the link between the legitimacy of decision-making and rule of law requirements such as acting within powers, due process, reason-giving, proportionality and fundamental rights.¹⁰⁶ In a treaty-based system, such requirements may be found in textual requirements and/or in general principles of law,¹⁰⁷ and national courts may be obliged to review domestic legal acts with reference to those treaty standards (which may have an extraterritorial effect).¹⁰⁸ This is paradigmatic of the role played by the national courts of the EU's member states, where an occasional pluralist resistance to the demands of the supranational order can be contrasted with a much fuller acceptance of EU law's foundational doctrines and general principles of law.¹⁰⁹ However, it has also been noted that national court acceptance of those principles – as well as those in some other treaty systems – has been aided by the fact that many of the principles have been borrowed from the prior experience of national administrative laws.¹¹⁰ Global administrative law in that way recognises an historical role for national conceptions of legality, as they will have helped to shape the legality principle that now exists in the supranational and international settings. A contemporary role for national conceptions is likewise envisaged in so far as they may interact with the legal norms of other orders as a result of the many indirect, or incidental, linkages within the global polity.¹¹¹

Returning to *Edwards*, the legality principle that is involved in the case assumes a more complex quality. Certainly, the account that would be offered by international law orthodoxy – that the EU and UK were merely acting in the light of their obligations under the Aarhus Convention – would understate the significance of common law (and European) developments that pre-dated the Aarhus Convention's ratification. The point here, again, is

103 Cases C-402 and 415/05 *Kadi v Council and Commission* [2008] ECR I-6351. Compare the reasoning – though not the outcome – in Case T-85/09 *Kadi v Commission and Council* [2011] 1 CMLR 24. Compare, too, the approach of the ECtHR in *Al-Jedda v UK* (2011) 53 EHRR 23.

104 T Tridimas and J A Gutierrez-Fons, 'EU Law, International Law, and Economic Sanctions against Terrorism: The Judiciary in Distress?' (2008) 32 *Fordham International Law Journal* 660.

105 On which see C Harlow, 'Accountability as a Value for Global Governance and Global Administrative Law' in Anthony et al (n 100) 173.

106 See generally Kingsbury et al (n 7).

107 J Crawford, *Brownlie's Principles of Public International Law* (8th edn, OUP 2012).

108 M Gondek, *The Reach of Human Rights in a Globalising World: Extraterritorial Application of Human Rights Treaties* (Hart Publishing 2009); and eg *Al-Skeini v UK* (2011) 53 EHRR 18.

109 On the instances of resistance see Craig and de Búrca (n 52) ch 9.

110 The seminal account is J Schwarze, *European Administrative Law* (2nd edn, Sweet & Maxwell 2006). See too P Birkinshaw, *European Public Law* (Butterworths 2003) ch 3; and, as regards the ECHR, D Leonardi, 'The Strasbourg System of Human Rights Protection: "Europeanisation" of the Law through the Confluence of the Western Legal Traditions' (1996) 8 *ERPL* 1139.

111 Auby (n 8); and J-A Scholte, *Globalisation: A Critical Introduction* (2nd edn, Palgrave Macmillan 2005).

that common law approaches to matters of standing, costs and interventions had already emphasised the public interest in vindicating the rule of law, where the relevant ‘rules of law’ had included EU measures in the field of the environment.¹¹² While the Aarhus Convention purported to embed liberal procedural regimes as a means for achieving more effective protection of the environment, its requirements arguably still did nothing more than complement well-established common law practices. On this view, the common law principle of legality became an active facilitator of international norms, rather than something that was subsumed by them.

The complexity of the legality principle can also be seen in *Ahmed v HM Treasury*.¹¹³ This was a fundamental rights case in which JUSTICE intervened as a third party on the question of how the common law should protect rights. The specific issue in the case was the legality of Orders in Council that had been made under section 1 of the United Nations Act 1946 and which, as with the disputed EU Regulations in *Kadi*, gave effect to UN Security Council Resolutions on the freezing of the financial assets of individuals involved in terrorism. According to section 1 of the Act, Orders in Council could make such provision as appeared ‘necessary or expedient’ for the purposes of implementing UN Resolutions, and the government had thereby made the Terrorism (United Nations Measures) Order 2006 (the Terrorism Order) and the Al-Qaida and Taliban (United Nations Measures) Order 2006 (the Al-Qaida Order).¹¹⁴ The Terrorism Order was duly quashed by the Supreme Court for the reason that it had included a test of ‘reasonable suspicion’ about an individual’s involvement in terrorism, which was held to have taken it beyond the wording of the corresponding UN Resolution and any argument that it had been ‘necessary or expedient’ to include the test. However, more remarkable was the Supreme Court’s ruling on the legality of the Al-Qaida Order, where article 3(1)(b) had been challenged as contrary both to the Human Rights Act 1998/ECHR and common law fundamental rights standards. Under article 3(1)(b), individuals whose names had been added to the UN sanctions list automatically had their assets frozen in the UK even though the UN listing process did not observe minimum fair-hearing guarantees such as the right of access to a court. While the Supreme Court rejected the argument that this rendered article 3(1)(b) unlawful under the Human Rights Act 1998 – the Court noted Article 103 of the UN Charter when holding that the ECHR was subject to the prior force of UN Resolutions¹¹⁵ – it held that art 3(1)(b) was contrary to the common law’s legality principle. When doing so, it referred to the interpretive rule whereby general words in an Act of Parliament cannot provide a basis for an interference with fundamental rights, as such interference can be achieved only where Parliament uses express words to that effect or words that have that effect by way of necessary implication.¹¹⁶ Noting that s 1 of the United Nations Act 1946 did not confer such a power upon the executive, the Supreme Court quashed article 3(1)(b) given its proximity to a procedural regime that was lacking minimum fair-hearing guarantees. The Supreme Court also refused a Treasury request for a suspension of the

112 *Eg R v Inspectorate of Pollution, ex p Greenpeace (No 2)* [1994] 4 All ER 329.

113 [2010] UKSC 2 and 5; [2010] 2 AC 534.

114 SI 2006/2657 and SI 2006/2952, respectively.

115 Following *R (Al-Jedda) v Secretary of State for Defence* [2007] UKHL 58, [2008] AC 332; and see *Al-Jedda v UK* (2011) 53 EHRR 23. Art 103 of the UN Charter provides: ‘In the event of a conflict between the obligations of the Members of the United Nations under the present Charter and their obligations under any other international agreement, their obligations under the present Charter shall prevail.’

116 See [2010] 2 AC 534, 626 [46], Lord Hope quoting *R v Secretary of State for the Home Department, ex p Pierson* [1998] AC 539, 575, Lord Browne-Wilkinson; and [2010] 2 AC 534, 646 [111], Lord Phillips quoting *R v Secretary of State for the Home Department, ex p Simms* [2000] 2 AC 115, 131, Lord Hoffmann.

remedies pending legislative amendment as that would ‘obfuscate the effect of its judgment’ by suggesting that the relevant provisions were in some way valid.¹¹⁷

Ahmed was inevitably a controversial ruling and Parliament soon enacted primary legislation that had the effect of overriding it.¹¹⁸ Nevertheless, the case still provides an insight into the robust nature of the common law’s legality principle and its protection of, in this instance, the right of access to a court.¹¹⁹ Although the Supreme Court chose not to situate its reasoning within the Human Rights Act 1998 and the analogical pluralism of the *Kadi* case – *Kadi* was read as specific to the EU legal order – its reliance on the common law principle of legality had essentially the same effect as the CJEU’s ruling. This thus marked a common law limitation to the flow of international norms and, even though the subsequent parliamentary intervention removed that limitation, the principle of legality apparently also countenances the review of an Act of the Westminster Parliament. This is the territory of the *Jackson* case in which Lord Steyn famously cautioned that: ‘[I]n exceptional circumstances involving an attempt to abolish judicial review or the ordinary role of the courts . . . [judges] may have to consider whether this is a constitutional fundamental which even a sovereign Parliament acting at the behest of a complaisant House of Commons cannot abolish.’¹²⁰ Marched to its logical conclusion, this would suggest that it is the common law principle of legality that lies at the very heart of the UK constitution, and that the principle’s global relevance should be seen in that light.

An equally controversial case – albeit for very different reasons – was *R (Corner House Research) v Director of the Serious Fraud Office*.¹²¹ The claimant was here given standing to challenge a decision of the Director whereby he had discontinued an investigation into allegations of corruption in BAE System’s dealings with officials in Saudi Arabia. The investigation had been commenced on the basis of the Director’s powers under section 1(3) of the Criminal Justice Act 1987 as read with sections 108–10 of the Anti-terrorism, Crime and Security Act 2001 (the latter provisions giving effect to the UK’s obligations under the Organisation for Economic Co-operation and Development (OECD) Convention on Combating Bribery of Foreign Public Officials in International Business Transactions 1997).¹²² As the investigation progressed, the Saudi authorities made an explicit threat that they would withdraw cooperation with the UK on matters of anti-terrorism, and there were very real concerns within the UK government that this would put ‘British lives on British streets’ at risk.¹²³ The Director’s decision to discontinue was therefore taken with reference to ‘the public interest’ in ‘public safety’, and *Corner House* challenged that decision with reference to the competing public interest in upholding the rule of law. The claimant’s arguments were to succeed in the Divisional Court, which held that the Director had acted unlawfully by, in effect, surrendering his discretion in the face of the Saudi threat. However, the House of Lords allowed the Director’s appeal for reasons associated with the separation of powers doctrine. Having noted the breadth of the discretion that is generally entrusted to public prosecutors under UK law, the House of Lords emphasised the limited role that

117 [2010] 2 AC 534, 690 [8], Lord Phillips.

118 The Terrorist Asset-Freezing Act (Temporary Provisions) Act 2010 and the Terrorist Asset-Freezing Act 2010.

119 See too *Al Rawi v Security Service* [2011] UKSC 34, [2012] 1 AC 531. Compare *Tariq v Home Office* [2011] UKSC 35, [2012] 1 AC 452. See also *Bank Mellat v HM Treasury* [2013] UKSC 38 and [2013] UKSC 39.

120 *R (Jackson) v Attorney-General* [2005] UKHL 56, [2006] 1 AC 262, 302–03 [102]. See also Baroness Hale’s comments at [2006] 1 AC 262, 318 [159].

121 [2008] UKHL 60, [2009] 1 AC 756.

122 See now the Bribery Act 2010.

123 [2008] UKHL 60, [2009] 1 AC 756, 835 [14], Lord Bingham quoting from the evidence.

the courts should play when decisions are challenged by way of judicial review.¹²⁴ The House also noted that the facts of the case had placed the Director in an ‘impossible situation’ and that he had been entitled, in law, to make the decision to discontinue.¹²⁵ As Lord Bingham expressed it:

The Director was confronted by an ugly and obviously unwelcome threat. He had to decide what, if anything, he should do . . . The issue in these proceedings is not whether his decision was right or wrong . . . but whether it was a decision which the Director was lawfully entitled to make . . . In the opinion of the House the Director’s decision was one he was lawfully entitled to make. It may indeed be doubted whether a responsible decision-maker could, on the facts before the Director, have decided otherwise.¹²⁶

There are perhaps two ways in which *Corner House* can be assessed. The first is to criticise it for equating the public interest in the rule of law with matters such as public safety, as this diminishes the prior normative force of the legality principle. This is certainly a view that has been advanced in some commentary on the case¹²⁷ and, cast in terms of global administrative law, it might be said that *Corner House* detracted from the value of accountability that underlies the OECD Convention on Bribery. However, a second view would hold that the restraint that characterised the House of Lords ruling is to be welcomed as it is as elemental to the legality principle as is judicial intervention in appropriate cases. The point here, again, is that public interest litigation has the potential to ‘politicise’ the courts by requiring them to adjudicate on matters that will involve difficult value judgments and the mobilisation of political knowledge.¹²⁸ On this more benign reading, *Corner House* amounts to nothing other than a reminder that global administrative law itself needs to accommodate a separation of powers doctrine.

Conclusion

This article began by noting that the term ‘the public interest’ occupies a central – if ill-defined – place in judicial review in the UK. Its corresponding analysis of public interest litigation across the three dimensions of judicial review has shown that UK courts now give effect to a legality principle that can both constrain decision makers within the UK and have implications for those outside it. While the genesis of the case law has been found very much in the common law’s approach to matters of standing, costs, and third-party interventions, it has been seen that the European and international law dimensions can also encourage – and, in some instances, demand – liberal access to the courts. This has resulted in cases coming before the courts where the points of law at issue have transcended the three dimensions of judicial review and required the courts to elaborate upon the requirements of legality in an era of global governance. When doing so, the courts have revealed much about the role that they might play in developing legal principles within an emerging body of global administrative law.

Of course, much of this presupposes the legitimacy of the link that the courts have made between ‘the public interest’ and the ‘rule of law’, and this is where criticisms of the case law may remain. Certainly, the fact that pressure groups and so on may associate the

124 [2008] UKHL 60, [2009] 1 AC 756, 840–41 [30–31], Lord Bingham citing, among others, *R v DPP, ex p Manning* [2001] QB 330; *R (Birmingham) v Director of the SFO* [2007] QB 727; and *Mobit v DPP of Mauritius* [2006] 1 WLR 3343.

125 [2008] UKHL 60, [2009] 1 AC 756, 848 [57], Baroness Hale.

126 [2008] UKHL 60, [2009] 1 AC 756, 844–5 [41–42].

127 J Jowell, ‘Caving in: Threats and the Rule of Law’ (2008) 13 JR 273.

128 Harlow (n 59).

public interest with particular policy preferences weakens the idea that public interest litigation will always be synonymous solely with the rule of law. Moreover, where pressure groups and so on rely upon European law and/or international law to challenge measures that may include Acts of the Westminster Parliament, this raises related concerns about how such norms are formulated and why they should prevail over deliberative choices that have been taken at the national level. This is largely a concern about the impact of elitism above the level of the state,¹²⁹ and it resonates with a much wider critique about the need for a revised, post-national model of democracy in a globalised era.¹³⁰ That said, global administrative law arguably offers some means for achieving accountability and control in that era, and it engages the UK courts both directly and indirectly in that endeavour. While there will always be scope for debate about whether judicial intervention is merited in a particular case – *Corner House* providing but one example – there can surely be no dispute about the imperative nature of the judicial review function. Any other perspective would only threaten the rule of law as a defining constitutional value.¹³¹

129 J Morison and G Anthony, 'The Place of Public Interest' in Anthony et al (n 100) 215.

130 For possibilities, see G De Búrca, 'Developing Democracy beyond the State' (2008) 46 *Columbia Journal of Transnational Law* 101.

131 R (*Jackson*) v *Attorney-General* [2005] UKHL 56, [2006] 1 AC 262, 304 [107], Lord Hope; and Constitutional Reform Act 2005, s 1.

Cartel criminalisation and due process: the challenge of imposing criminal sanctions alongside administrative sanctions within the EU

DR PETER WHELAN*

University of East Anglia

Abstract

There is increasing debate within the European Union (EU) concerning the imposition of criminal sanctions upon those individuals who engage in cartel activity. For it to be legitimate, such cartel criminalisation must respect the due process guarantees contained in the European Convention on Human Rights (ECHR). Unfortunately the literature on this issue is deficient and the specifics of this legal challenge are not fully understood. In particular, a comprehensive analysis of the due process-related challenge presented when personal criminal antitrust sanctions are employed alongside administrative sanctions for a given cartel is conspicuously absent from the literature. This article rectifies this deficiency by examining this particular legal challenge and its relevance to information exchange, double jeopardy and concurrent antitrust proceedings. In doing so, it identifies practical techniques designed to meet the challenge of due process in this context, as well as the inherent tensions between due process and the objectives of European antitrust criminalisation.

Keywords: anti-cartel enforcement; concurrent proceedings; criminal cartel sanctions; due process; EU competition law; information exchange; *ne bis in idem*.

Introduction

‘Cartel activity’ encapsulates the making or implementing of an anticompetitive agreement, concerted practice or arrangement by competitors to fix prices, rig bids, establish output restrictions or divide markets.¹ Such activity is prohibited by all of the national competition laws of the EU member states. Furthermore, if it affects trade between the member states, cartel activity is prohibited by art 101(1) of the Treaty on the Functioning of the European Union (TFEU), which is enforced by the European Commission (the Commission) and the national competition authorities (NCAs) and courts of the member states.² Traditionally, within Europe, cartel law enforcement has tended to

* PhD (Cantab), Senior Lecturer in Law, UEA Law School and ESRC Centre for Competition Policy, University of East Anglia; email: p.wheLAN@uea.ac.uk.

1 OECD, *Recommendation of the OECD Council Concerning Effective Action Against Hard Core Cartels*, 25 March 1998, [2(a)].

2 See Council of Ministers, Council Regulation (EC) No 1/2003 of 16 December 2002 on the implementation of the rules on competition laid down in arts 81 and 82 of the Treaty, [2003] OJ L1/1 (hereafter ‘reg 1/2003’).

avoid the employment of personal criminal punishment.³ Recently, however, there has been increasing debate concerning the imprisonment of individual cartelists.⁴ This debate is more than an academic exercise: some European countries have criminalised their antitrust regimes, while others have contemplated doing so.

Cartel criminalisation within the EU presents complex challenges. Advocates of criminalisation must demonstrate, for example, that it is justifiable in principle. This is often accomplished using (economic) deterrence or retribution theory; deterrence and retribution therefore represent the primary enforcement objectives of European antitrust criminalisation.⁵ Additionally, advocates must demonstrate that such criminalisation can adhere to relevant mandatory legalities. In particular, to attain legitimacy, antitrust criminalisation should respect the due process guarantees contained in the ECHR. It is vital therefore to understand how due process impacts upon the mechanics of antitrust criminalisation and whether practical measures and techniques are required to respect due process. Importantly, one must also understand whether respecting due process impacts negatively upon the two primary objectives of European antitrust criminalisation. Indeed, the existence of such an impact should influence the initial decision whether to criminalise. Failure to respect due process in this context may lead to the creation of a criminal antitrust regime which is not only legally unsound, but which also wastes resources and creates negative publicity for the antitrust regime.

Admittedly, one aspect of due process has been considered in the literature: the claim that antitrust criminalisation inevitably results in a 'strengthening of rights' to the benefit of the accused and to the detriment of the relevant enforcement objectives.⁶ In examining the 'strengthening of rights' contention, one analyses whether imposing criminal sanctions *rather than* administrative sanctions for a given cartel engenders potential negative implications for antitrust enforcement due to human rights considerations. However, to present a complete picture of the challenge of due process for European antitrust criminalisation, one must go further and consider the potential interaction of both criminal and administrative cartel sanctions in this context. Specifically, one should first explore whether the introduction of criminal antitrust sanctions does not preclude the imposition of administrative sanctions *alongside* such criminal sanctions *for a given cartel* (herein termed the 'additional dynamic' contention), before examining the impact of that particular contention upon criminal antitrust enforcement that respects due process.

The current (fragmented) literature alludes somewhat to the additional dynamic contention and its potential impact. Perrin, for example, argues that, due to concerns for human rights, the fact that criminal antitrust sanctions exist in some EU member states but not in others (where only administrative sanctions exist) impacts negatively on European cooperation in antitrust enforcement.⁷ Zuleeg questions whether the mere existence of both criminal and administrative antitrust procedures within Europe has the potential to

3 See C Harding, 'Business Collusion as a Criminological Phenomenon: Exploring the Global Criminalisation of Business Cartels' (2006) 14 *Critical Criminology* 181.

4 See generally C Beaton-Wells and A Ezrachi (eds), *Criminalising Cartels: Critical Studies of an International Regulatory Movement* (Hart Publishing 2011).

5 See P Whelan, *The Criminalization of European Cartel Enforcement: Theoretical, Legal, and Practical Challenges* (OUP forthcoming).

6 P Whelan, 'Criminal Cartel Enforcement in the European Union: Avoiding a Human Rights Trade-Off' in Beaton-Wells and Ezrachi (n 4).

7 B Perrin, 'Challenges Facing the EU Network of Competition Authorities: Insights from a Comparative Criminal Law Perspective' (2006) 31(4) *European Law Review* 540.

violate the accused's right not to be punished twice for the same conduct.⁸ Finally, Nazzini considers whether the existence of the UK Cartel Offence introduces human rights-based problems due to the possibility of concurrent criminal and administrative proceedings.⁹ Notwithstanding this literature, to date, no systematic analysis has been undertaken concerning the validity and impact of the additional dynamic contention upon European cartel criminalisation. This situation is regrettable.

This article examines the validity of the additional dynamic contention, as well as its impact upon a project of European antitrust criminalisation which respects due process. It is divided into two substantive sections. Section 1 determines the *validity* of the additional dynamic contention by considering the likelihood and legality of four of its potential manifestations. By contrast, section 2 examines the *impact* of the additional dynamic contention: it examines the specifics of the due process-related challenge presented when personal criminal antitrust sanctions are indeed employed alongside administrative sanctions. Three issues are considered: information exchange; double jeopardy; and concurrent proceedings. Practical techniques designed to meet the challenge of respecting due process are identified, as are the inherent tensions between due process and the objectives of European antitrust criminalisation, before concluding observations are offered.

1 Determining the validity of the additional dynamic contention

The additional dynamic contention posits that the introduction of criminal antitrust sanctions in a particular jurisdiction does not preclude the imposition of administrative sanctions *alongside* such criminal sanctions *for a given cartel*. There are four potential manifestations of this contention within the EU: (a) a national regime imposes both administrative sanctions on undertakings and criminal sanctions on individuals; (b) a national regime imposes criminal sanctions on individuals and a different national regime imposes administrative sanctions on undertakings; (c) the Commission imposes administrative sanctions on undertakings and a national regime imposes criminal sanctions on individuals; and (d) the Commission imposes criminal sanctions on individuals and a national regime imposes administrative sanctions on undertakings. To examine the *validity* of the additional dynamic contention, one must therefore examine the *legality and likelihood* of manifestations (a) to (d).

(A) INQUIRIES RELEVANT TO THE DETERMINATION OF VALIDITY

Three inquiries should be undertaken to examine the legality and likelihood of the identified manifestations of the additional dynamic contention: (i) whether EU law allows a national regime to adopt both criminal and administrative sanctions; (ii) whether the EU can mandate the adoption of (and/or impose) criminal antitrust sanctions; and (iii) whether diversity concerning antitrust sanctions actually exists across the member states.

(i) National competences concerning antitrust sanctions

EU law does not prevent member states from choosing whether to impose criminal and/or administrative sanctions for violation of art 101 TFEU. This is clear from reg 1/2003. Recital 16 of that particular regulation, for example, acknowledges the possibility of diversity in enforcement when it provides that 'as regards natural persons, they may be

8 M Zuleeg, 'Criminal Sanctions to be Imposed on Individuals as Enforcement Instruments in European Competition Law' in C D Ehlermann and I Atanasiu (eds), *European Competition Law Annual 2001: Effective Private Enforcement of EC Antitrust Law* (Hart Publishing 2002) 13.

9 R Nazzini, 'Criminalisation of Cartels and Concurrent Proceedings' (2003) 24(10) *European Competition Law Review* 483.

subject to substantially different types of sanctions across the various systems'. Furthermore, art 5 of the same regulation explicitly grants the member states the right to decide on the nature of the antitrust sanction to be imposed for violation of art 101 TFEU: the NCAs may take decisions 'imposing fines, periodic penalty payments or *any other penalty* provided for in their national law'.¹⁰ It is clear from art 12(3) of reg 1/2003 that the 'other penalty' referred to includes individual criminal sanctions.¹¹ The essence of art 12(3) is that the NCAs may exchange information for the purpose of enforcing art 101 TFEU, but such information may only be used in evidence to impose *custodial sanctions* on individuals when both the sending and the receiving member states impose 'sanctions of a similar kind'. Nothing in the regulation or in EU jurisprudence prevents the use of both administrative and criminal sanctions by national regimes for a given cartel. Regulation 1/2003 reflects the traditional deference of the European institutions to national procedural autonomy. Consequently, and subject to general principles of EU law, such as equivalence and effectiveness, the member states have the freedom to choose their type of antitrust sanctions. Under EU law at least, both criminal and administrative sanctions remain possible sanctions for cartel activity.

Even if EU law does not prohibit different types of antitrust sanctions, it does not necessarily follow that national law will be so flexible. In some countries particular types of sanctions may be prohibited by national constitutions. In Ireland, for example, sanctions imposed for violation of art 101 TFEU cannot be administrative sanctions: due to arts 34(1) and 37(1) of the Irish Constitution, antitrust sanctions can only be imposed by courts composed of judges appointed in accordance with the Irish Constitution.¹² This situation is not the norm across Europe, however. National regimes usually have the choice between criminal and administrative antitrust sanctions. Accordingly, subject to the (Irish-inspired) caveat above, '[t]he national systems to enforce European competition rules can be designed as a criminal law system, an administrative law system or both'.¹³

(ii) EU competence concerning criminal antitrust sanctions

In 2005 Wils argued that, under the (now defunct) European Community (EC) treaty framework, the European institutions could both mandate the introduction of criminal antitrust sanctions in the member states and impose such sanctions on individuals at the level of the (then) EC institutions.¹⁴ He argued: (a) that criminal EC antitrust sanctions were not expressly *prohibited* by EU law; (b) that the Commission had broad legislative enforcement powers under art 83 of the Treaty establishing the European Community (TEC) to adopt 'the appropriate regulations or directives to give effect to the principles set out in [art] 81 [TEC]', principles that underlined the undesirability of hard-core cartels; and (c) that art 81 TEC's use of the term 'undertaking' did not preclude individual (criminal) sanctions, as the measures adopted under art 83 TEC gave effect to the *principles* inherent in art 81 TEC. Wils acknowledged, however, that his argument would be undermined if the EU Constitution entered into force: the argument relied upon (then) art 47 of the Treaty on European Union (TEU) and the fact that the EU Treaty provisions on criminal matters had no impact upon the TEC antitrust provisions; the EU Constitution would have integrated the areas covered

10 Emphasis added.

11 See W Wils, 'Is Criminalization of EU Competition Law the Answer?' (2005) 28(2) *World Competition* 17.

12 See P Massey, 'Criminal Sanctions for Competition Law: A Review of Irish Experience' (2004) 1(1) *Competition Law Review* 23, 24.

13 O Jansen, 'The Systems of International Cooperation in Administrative and Criminal Matters in Relation to Regulation EC 1/2003' in G Dannecker and O Jansen (eds), *Competition Law Sanctioning in the European Union* (Kluwer Law International 2004) 258.

14 Wils (n 11).

by the TEC and by the other Titles of the EU Treaty into 'a single Treaty, without any hierarchy of the former over the latter', ensuring that the antitrust provisions would no longer be treated separately from (limited) EU powers over criminal matters.¹⁵

An identical effect occurred with the adoption of the Lisbon Treaty.¹⁶ Although the Lisbon Treaty actually consists of two founding treaties, it abolished any hierarchy between the amended TEU and the TEC (renamed the TFEU). Indeed, the Lisbon Treaty established one single union legal order founded on two treaties with the same value.¹⁷ Consequently, the antitrust provisions must be interpreted in light of those provisions concerning judicial cooperation on criminal matters. Union legislative competence in the criminal sphere is now provided for in the context of judicial cooperation on criminal matters.¹⁸ Consequently, one cannot argue that art 103 TFEU (the provision in the TFEU that is equivalent to art 83 TEC) can be used to legislate for criminal sanctions. Regarding the TFEU provisions on judicial cooperation on criminal matters, it is stipulated that in a certain number of enumerated areas directives may be adopted establishing minimum rules concerning the definition of criminal offences and sanctions. Cartel activity is not mentioned but may be included by virtue of art 83(1), para 3, TFEU. For this to happen, the European Council must unanimously adopt a decision after obtaining the consent of the European Parliament. Legislative competence in criminal matters is restricted to the use of directives and therefore would not cover the criminalisation of cartel activity at the level of the EU institutions. So antitrust criminalisation via the EU institutions is facilitated by the Lisbon Treaty, but it is confined to harmonisation of the enforcement regimes of the member states and is possible only if a consensus for such sanctions exists among the member states.

(iii) Diversity of antitrust sanctions across the EU

Currently there exists 'a wide variation in the nature and the level of the penalties' which can be imposed by the Commission and by the NCAs.¹⁹ While criminal antitrust sanctions are present in some member states (such as Ireland, the UK, Estonia and Germany, for example), they are noticeably absent in many others (including, for example, Spain, Sweden, Luxembourg and Holland), where administrative sanctions are the norm.²⁰ Furthermore, within the criminalised jurisdictions one often finds non-criminal antitrust sanctions which are aimed at undertakings. Such is the case with the UK and Germany, for example. Thus, there is indeed a diversified range of antitrust sanctions, both across the EU and within (some of) the member states.

This diversity is likely to continue. First, for efficiency reasons, criminal antitrust sanctions are unlikely to replace (as opposed to complement) administrative sanctions within a given regime.²¹ Second, not all of the legislative bodies of the member states may be convinced of the merits of European antitrust criminalisation. Such criminalisation presents complex theoretical, legal and practical challenges which may be perceived by a given jurisdiction to be too demanding to be taken on at present. Third, different antitrust

15 Wils (n 11) 158.

16 See P Whelan, 'Contemplating the Future: Personal Criminal Sanctions for Infringement of EC Competition Law' (2008) 19(2) *King's Law Journal* 364.

17 See arts 1(3) and 1(2) TFEU.

18 See arts 82–86 TFEU.

19 W Wils, 'Community Report' in D Cahill and J D Cooke (eds) (2004), *The Modernisation of EU Competition Law Enforcement: FIDE 2004 National Reports* (CUP 2004) [190].

20 See Perrin (n 7) 552–54; and M O'Kane, *The Law of Criminal Cartels: Practice and Procedure* (OUP 2009) 325–27.

21 See eg OFT, *An Assessment of Discretionary Penalties Regimes* (OFT 1132, October 2009) [1.4].

authorities will have different priorities. While some member states have a mature antitrust regime and are better placed to consider criminalisation, others have only been enforcing competition provisions for a short period of time. The latter jurisdictions may (wisely) wish to develop a ‘competition culture’ before experimenting with criminal sanctions.²² Indeed, a strong competition culture and a political commitment to antitrust law are essential for effective criminal antitrust enforcement. Fourth, diversity should continue to exist not only across the EU but also within a given member state. For reasons of deterrence, criminalised jurisdictions should continue to impose administrative cartel sanctions on undertakings: by maintaining the administrative offences, they ensure that firms would not be incentivised to encourage cartel activity among their employees.²³ Finally, some countries may have more pronounced institutional challenges concerning antitrust criminalisation than others. Sweden, for example, would have to confront its legal prohibition on prosecutorial discretion to ensure that criminal antitrust sanctions do not undermine administrative leniency programmes.²⁴ Such institutional differences reduce the attractiveness of criminal sanctions for some jurisdictions. Provided that other member states remain committed to antitrust criminalisation, this fact alone ensures continued diversity across the EU.

(B) APPLICATION OF FINDINGS TO THE ‘FOUR MANIFESTATIONS’

According to the above findings, only one out of the four manifestations is invalid. The invalid manifestation is the following: the Commission imposes criminal antitrust sanctions on individuals and a national regime imposes administrative antitrust sanctions on undertakings. It is clear from the analysis concerning EU criminal competence that the European institutions cannot actually impose criminal antitrust sanctions, even if, provided strict conditions are fulfilled, they may mandate their adoption within the member states. The three remaining manifestations are valid, however.

The first manifestation (that is, where a given national regime imposes both administrative and criminal sanctions) occurs within the UK, for example, and is not prohibited under EU law. National law may prohibit administrative sanctions, but this will not be the case in every member state. The second manifestation (that is, where a national regime imposes criminal antitrust sanctions on individuals and a different national regime imposes administrative sanctions on undertakings) is facilitated by art 5 of reg 1/2003 and is likely given the different priorities facing different member states and the current absence of a European consensus about the desirability of antitrust criminalisation. The third manifestation (that is, where the Commission imposes administrative antitrust sanctions on undertakings and a national regime imposes individual criminal sanctions) is also valid. This is due, again, to art 5 of reg 1/2003, as well as to the fact that the Commission cannot be given the power to impose criminal sanctions. Hence, given that three out of its four manifestations are both legal and likely, the additional dynamic contention itself is valid.

2 Determining the impact of the additional dynamic contention

This section acknowledges the validity of the additional dynamic contention and examines its actual *impact* upon European antitrust criminalisation. Three important issues are considered: (a) information exchange; (b) double jeopardy; and (c) concurrent proceedings.

22 See J Fingleton, M B Girard and S Williams, ‘The Fight Against Cartels: Is a “Mixed” Approach to Enforcement the Answer?’ in B Hawke (ed), *International Antitrust Law and Policy: Fordham Corporate Law 2006* (Juris Publishing 2007) 22.

23 See P Whelan, ‘A Principled Argument for Personal Criminal Sanctions as Punishment under EC Cartel Law’ (2007) 4(1) *Competition Law Review* 7, 30–31.

24 See Swedish Competition Authority, ‘Criminal Enforcement of Antitrust Laws’ (Fordham Law Seminar, New York, 14 September 2006).

(A) EXCHANGE OF INFORMATION

The effectiveness of an antitrust investigation often depends upon the extent of cooperation with and assistance received from foreign states.²⁵ This is particularly so for cartels: they often cross the boundaries of jurisdictions and ‘an agency may not be aware of a cartel affecting its jurisdiction, while another agency has cognisance of it’.²⁶ As cartels become increasingly international in nature, cooperation in cartel cases will continue to grow in importance.²⁷ Cooperation may take many forms, such as the exchange of information and the collection of evidence. There is consensus on the importance of cooperation in the form of information exchange.²⁸ Information exchange helps, *inter alia*, to increase rates of detection and has a positive impact upon the enforcement objectives of antitrust law. Such cooperation is therefore valuable and should be facilitated where appropriate.

A due process problem may arise with information exchange (within the second and third manifestations of the additional dynamic contention) when the following conditions are fulfilled: (a) personal criminal sanctions are imposed in one jurisdiction for cartel activity, and administrative sanctions are imposed in another; (b) both jurisdictions are investigating a given cartel; and (c) the criminal jurisdiction requests from the other jurisdiction information that has been collected during its respective investigation. If information collected during the administrative investigation could legally be provided to the other jurisdiction (and that jurisdiction could use it in evidence in a criminal trial), potential would exist for a violation of due process: the collection of evidence in the requested state will presumably have been conducted according to administrative standards and not criminal standards. This potential problem should be neutralised for European antitrust criminalisation to be legally sound.

(i) Regulation 1/2003 and the European Competition Network

Due to reg 1/2003, the Commission and the NCAs cooperate with one another through the European Competition Network (ECN).²⁹ This network is an official forum for discussion and cooperation in the application and enforcement of EU competition policy.³⁰ For the purpose of applying arts 101 and 102 TFEU, the Commission and the NCAs have the power to provide one another with and use in evidence any matter of fact or of law, including confidential information.³¹ Such cooperation, while legal, is not mandatory. There is a narrow exception to the rule that the information exchanged should only be used in evidence for the purpose of applying arts 101 and 102 TFEU and in respect of the subject-matter for which it was collected by the transmitting authority:³² information exchanged under art 12 may also be used for the application of national competition law, but only where that law ‘is applied in the same case and in parallel to Community competition law

25 Perrin (n 7) 540.

26 ICN, ‘Co-operation between Competition Agencies in Cartel Investigations’ (Report to the ICN Annual Conference, Moscow, 7 May 2007).

27 ICN (n 26) 5.

28 See M Bloom, ‘Exchange of Confidential Information among Members of the EU Network of Competition Authorities’ in C D Ehlermann and I Atanasiu (eds), *European Competition Law Annual 2002* (Hart Publishing 2004) 389.

29 See reg 1/2001, recital 15 and ch IV (in particular arts 11 and 12).

30 See European Commission, *Commission Notice on Cooperation within the Network of Competition Authorities*, 2004/C 101/03, Brussels, 24 April 2004 [1].

31 Reg 1/2003, art 12(1). See also arts 11(4) and 22(1).

32 See reg 1/2003, art 12(2); and *Dow Benelux v Commission* (85/87) [1989] ECR 3137.

and does not lead to a different outcome'.³³ This power cannot be undermined by national law.³⁴ Pursuant to this power, the exchange of information can be either vertical (for instance, from the Commission to the NCAs) or horizontal (that is, between the NCAs).³⁵ Consequently, the power to provide information facilitates three types of information flows: from an administrative regime to a criminal regime; from an administrative regime to another administrative regime; and from a criminal regime to an administrative regime.

(ii) Neutralising potential problems of due process with criminalisation within the ECN

For present purposes, it is the first information flow that is potentially problematic. If during an administrative investigation the Commission discovered evidence of an individual's participation in a cartel involving harm to UK consumers, for example, and was allowed to pass that information on to the UK authorities for them to use as evidence in a criminal court against that individual, the collection of (some of the) evidence would be conducted according to administrative standards and not criminal standards. Administrative protections may be less robust than those required in a criminal context and are in fact so with the Commission's investigations.³⁶ Therefore if the power to use exchanged information in this context were *unlimited*, due process would be undermined.³⁷ To avoid this, reg 1/2003 contains art 12(3).

1 Article 12(3) of reg 1/2003

Article 12(3) of reg 1/2003 states that information exchanged via the ECN can only be used in evidence to impose sanctions upon individuals where:

- the law of the transmitting authority foresees sanctions of a similar kind in relation to an infringement of art 101 TFEU or art 102 TFEU or, in the absence thereof,
- the information has been collected in a way which respects the same level of protection of the rights of defence of natural persons as provided for under the national rules of the receiving authority. However, in this case, the information exchanged cannot be used by the receiving authority to impose custodial sanctions.

Accordingly, reg 1/2003 accommodates antitrust criminalisation in the member states. Those member states which have 'sanctions of a similar kind'³⁸ (for example, custodial sanctions) are allowed to exchange amongst themselves and, importantly, use in evidence against individuals the information that they have collected. The presumption here is that, if personal criminal sanctions are imposed in one country, then due to, *inter alia*, art 6 ECHR, that country must conduct its investigations in a manner which respects human rights. Indeed, if both jurisdictions have custodial sanctions for cartel activity then the 'procedural safeguards in both systems are considered to be equivalent'.³⁹ While a 'common' minimum level of protection may well exist, it is also true that some human

33 Reg 1/2003, art 12(1). If member states wish to exchange information where these conditions are not met, they may require additional legislative provisions: Jansen (n 13) 259.

34 See European Commission (n 30) [27].

35 Jansen (n 13) 258; and European Commission (n 30) [27].

36 See Whelan (n 6).

37 See W Wils, *Efficiency and Justice in European Antitrust Enforcement* (Hart Publishing 2008) 194; and W Wils, *Principles of European Antitrust Enforcement* (Hart Publishing 2005) s 1.2.12.

38 It appears that the fundamental criterion that should be used to determine if the personal sanctions are of a 'similar kind' is whether the sanctions are custodial or non-custodial; see European Commission (n 30) [28(c)]; and L O Blanco (ed), *EC Competition Procedure* (2nd edn, OUP 2006) [3.30].

39 European Commission (n 30) [28(c)].

rights protections may be stronger in one country (due to its national law) than in another, even if both adhere to the ECHR requirements. Assume that both requested and requesting state have criminal antitrust sanctions. If the requested state goes no further than the requirements of art 6 ECHR in its protection of the rights of the accused, then, under reg 1/2003, it is that level of protection that is achieved. However, if the requested state has a level of protection that is more robust than that required by art 6 ECHR, the higher level of protection is operative. When both jurisdictions impose criminal sanctions, it is the human rights protections of the requested state that matter: it is irrelevant that the requesting state has in place higher levels of protection than the requested state. Due to the operation of art 12(3), then, *at the very least* (and assuming that the member states respect their ECHR obligations) art 6 ECHR will be respected when both jurisdictions impose criminal sanctions on natural persons; but more robust protections may also be secured, depending upon the (national) human rights law of the requested state.

A relatively similar situation is reached where only the requesting state has criminalised its regime: when the transmitting country does not have criminal sanctions and therefore has not necessarily gathered evidence according to criminal standards, such information can only be used as intelligence, not evidence, unless the human rights protections of the receiving country have been respected. It is assumed here that the human rights guarantees of the receiving country will *at the very least* respect art 6 ECHR. Again, the actual level of protection secured may be higher, and all depends upon the relevant national law. But this time, of course, the relevant national law is that of the requesting state, not the requested state. There is, however, an extra protection if the requested state does not employ criminal sanctions: if information is used in evidence in the requesting state, only non-custodial punishment may be imposed. These stricter rules are ostensibly designed to preserve the higher level of protection that individuals usually benefit from in comparison to undertakings.⁴⁰

2 Critically evaluating art 12(3)

One could argue that, as a direct result of art 12(3), the due process rights of the accused are adequately protected while: (a) no obstacles are placed in the way of information exchange between criminal antitrust jurisdictions; and (b) no unreasonable obstacles are placed in the way of information exchange between administrative and criminal antitrust jurisdictions. For this to be so, one must first accept the logic of the assumption that art 6 ECHR (or its equivalent) will be protected in the member states. This is not difficult to do. First, the ECHR, as an international treaty, is binding on all the EU member states⁴¹ and, although infringements of art 6 ECHR may occasionally occur in practice, the member states must nonetheless acknowledge and respect the mandatory nature of rulings addressed to them by the European Court of Human Rights (ECtHR).⁴² Member states, then, are restrained in their activities by the ECHR.⁴³ Second, fundamental rights are respected as general principles of EU law, and when member states apply EU law they are under the obligation to comply with EU requirements concerning the protection of fundamental rights. Such human rights protections find their inspiration in, *inter alia*, the ECHR, including art 6 thereof.⁴⁴ Consequently, the assumption appears to be sound. If so,

40 A Türk, 'Modernisation of EC Antitrust Enforcement' in H Hofmann and A Türk (eds), *EU Administrative Governance* (Edward Elgar Publishing 2006) 223.

41 D Gomien, *Short Guide to the European Convention on Human Rights* (Council of Europe 2005) 12.

42 See eg art 46(1) ECHR.

43 But this is not to say that the signatories must give direct effect to the ECHR within national law: *Ireland v UK* (1978) 2 EHRR 25. In fact, there is 'wide variation' in the extent to which the signatories have done so: M Janis, R Kay and A Bradley, *European Human Rights Law: Text and Materials* (3rd edn, OUP 2008) 850.

44 See art 6(3) TEU.

the exchange of information between two criminal antitrust jurisdictions is not problematic under European human rights law.

The case where information is exchanged between an administrative and a criminal regime is less straightforward, however. Again, the assumption that art 6 ECHR will be protected in the member states is made; and again that assumption appears to be sound. Hence, the human rights of the accused are protected (at least) to the ECHR standard. However, an extra obstacle exists concerning the use in evidence of the information, even where the art 6 ECHR rights of the accused have been protected: custodial sanctions cannot be imposed. It is not evident why such a rule (the 'custodial rule') is necessary. If the human rights protections of the requesting state are met then, putting aside the issue of proof concerning such protection, there should be no real issue concerning an imbalance of the requirements of due process between the administrative regime which gathered the information and the criminal one which receives it and uses it in evidence.

Perhaps human rights concerns are not at issue here and in fact politics dominates: some member states may have been concerned about the use of criminal sanctions per se to enforce EU competition law and may not have wished to see information collected in their respective jurisdictions being used to imprison executives in other member states. The existence of the custodial rule arguably illustrates that 'custodial sanctions for violations of competition law are not a matter of consensus within the Union at present'.⁴⁵ Working documents from the Council of Ministers support this view.⁴⁶ Nonetheless, although one can interpret art 12(3) as reflecting 'hostility' towards custodial sanctions,⁴⁷ such hostility is less openly reflected in reg 1/2003 than in the original Commission proposal for a new implementing regulation, where it was proposed 'to exclude the use of evidence exchanged for imposing penalties other than pecuniary penalties'.⁴⁸ Provided that the assumption about art 6 ECHR is sound, there is no reason founded in European human rights law for such a strict approach to the issue of information exchange. If reg 1/2003 is to be reformed,⁴⁹ then the (unnecessary) custodial rule should be abolished in the process: to the extent that the threat of imprisonment is more effective in ensuring deterrence than administrative/criminal fines,⁵⁰ such a rule reduces in practice the deterrent effect of (criminal) antitrust sanctions, without any justification based upon legality or practicality. The Commission itself has acknowledged that discussions have arisen concerning whether the custodial rule is 'is too far-reaching and is an obstacle to efficient enforcement'.⁵¹ Consequently, it 'may be appropriate to examine whether other options are available, while fully preserving parties' rights of defence'.⁵² One possible solution is the abolition of the custodial rule and the use of a 'double barrier' which allows for the information exchanged to be used in evidence to secure custodial sentences provided that: (a) the transmitting

45 C Gauer, 'Due Process in the Face of Divergent National Procedures and Sanctions' (Conference Paper, Antitrust Reform in Europe: A Year in Practice Papers, Brussels, 9–11 March 2005) 51.

46 See eg Document 11791/02 of 4 September 2002, which was produced by the Competition Working Party to the Permanent Representatives Committee: <<http://register.consilium.eu.int>> accessed 12 June 2013.

47 Perrin (n 7) 547.

48 European Commission, *Proposal for a Council Regulation on the Implementation of the Rules on Competition laid down in Articles 81 and 82 of the Treaty and Amending Regulations (EEC) No 1017/68, (EEC) No 2988/74, (EEC) No 4056/86 and (EEC) No 3975/87* (COM(2000) 582 final, Brussels, 27 September 2000) 21. See also *ibid* 41.

49 On this, see <http://ec.europa.eu/competition/consultations/2008_regulation_1_2003/index.html> accessed 12 June 2013.

50 See Wils (n 11); and Whelan (n 23).

51 European Commission, *Communication from the Commission to the European Parliament and the Council: Report on the Functioning of Regulation 1/2003* (COM(2009) 206 final, Brussels, 29 April 2009) [32].

52 European Commission (n 51) [32].

authority verifies that the information transmitted conforms with its rules on admissibility; and (ii) the receiving authority verifies that the information received can be used in conformity with its rules.⁵³ Such an approach would resolve the due process-related problem noted above, while minimising the negative impact upon the efficiency and effectiveness of European antitrust enforcement.

Interestingly, it has been argued that the *mere coexistence* of criminal and administrative sanctions inevitably places a major obstacle in the way of information exchange within the ECN. For Perrin, art 12(3) of reg 1/2003 results in ‘a considerable fracture in the supposedly unified EU Network into two distinct networks whose *fault line* is based on the decision of member states to impose custodial [antitrust] sanctions’.⁵⁴ Accordingly, EU jurisdictions imposing custodial sanctions for antitrust offences ‘will find themselves *excluded* from the larger EU Network and demoted to the smaller EU Network of like-minded member states’.⁵⁵ The author respectfully disagrees with this argument, for two reasons. First, the art 12(3) prohibition on the use of exchanged information to secure custodial sanctions in the *receiving* state only extends to the use by the receiving member state of such information *in evidence*. Member states may legally send information to other member states (which impose custodial sanctions) to facilitate the latter in *their investigation* of alleged cartel activity. Furthermore, the receiving member states (who impose custodial sanctions) are legally entitled to use such information to detect and to obtain proof of — rather than to use as evidence of — a violation of art 101(1) TFEU. Second, there is no formal/informal punishment mechanism which can be used against those member states without criminal antitrust sanctions who transfer information to member states with criminal sanctions where the latter then use that information in evidence to secure custodial sentences. In short, there is no negative ‘fallout’ for the sending member state. All that art 12(3) of reg 1/2003 does is ensure that the receiving member state will be restrained in its employment of the information: it does not ensure that such information will not be exchanged in the first place. If information is used in a manner which is inconsistent with art 12(3) of reg 1/2003, it is only the *receiving* member state that will be at fault. Consequently, it is not exactly clear why the mere exchange of information through the ECN will result in a two-tier entity.

Admittedly, the coexistence of criminal and administrative sanctions is not without its problems, even with art 12(3) of reg 1/2003. One particular difficulty comes to mind: determining whether the collection of information respects ‘the same level of protection of the rights of defence of natural persons as provided for under the national rules of the receiving authority’.⁵⁶ This inquiry is relevant when a member state that imposes administrative sanctions sends information to another member state which is then used in evidence to secure a criminal cartel conviction. In such a situation, the defendant could claim that such information should not be used in evidence against her as its collection did not respect the same level of rights protection provided to natural persons in the receiving state. Indeed, for Perrin, the use of information in such a context would ‘be highly vulnerable to frequent challenges on the part of defendants who are natural persons’.⁵⁷ This is a strong argument: criminal sanctions can be severe; and defendants will likely present

53 See European Commission, *Commission Staff Working Paper Accompanying the Communication from the Commission to the European Parliament and the Council: Report on the Functioning of Regulation 1/2003* (SEC(2009) 574 final, Brussels, 29 April 2009) [245].

54 Perrin (n 7) 555.

55 *Ibid.*

56 Reg 1/2003, art 12(3).

57 Perrin (n 7) 555.

whatever (reasonable) claim they can to avoid conviction. Nonetheless, irrespective of whether such claims by the defendant would be as frequent as contended by Perrin, it is true that, to rule on such a claim, the relevant national judge(s) would need to evaluate the protections secured in the sending state, an exercise that might be burdened by language difficulties and/or a lack of cooperation from the sending NCA. Reliable information on the procedural guarantees of other member states thus increases in importance due to European antitrust criminalisation. Obtaining such information in any given case increases the costs associated with enforcement. Challenges by defendants on this ground alone therefore have potential to affect the cost–benefit analyses that need to be undertaken by the authorities when deciding whether to introduce criminal antitrust sanctions for the purpose of deterrence: such (likely) challenges impact upon the efficiency of criminal antitrust enforcement.

(B) *NE BIS IN IDEM*

The principle of *ne bis in idem*, contained in art 4(1) of Protocol No 7 to the ECHR,⁵⁸ prevents an individual from being tried or punished more than once for the same offence in criminal proceedings under the jurisdiction of the same state. The principle only relates to a right not to be tried or punished twice in ‘criminal proceedings’.⁵⁹ However, under the ECHR the term ‘criminal’ is given an autonomous meaning⁶⁰ and, following the reasoning in *Menarini*,⁶¹ the term would cover (administrative) competition proceedings, such as those conducted at EU level by the Commission. This fact alone ensures that the principle remains relevant to the additional dynamic contention.

(i) The general operation of the principle

The guarantee enshrined in art 4 prohibits ‘trial and punishment “again” for an offence for which the applicant has already been finally acquitted or convicted’.⁶² Unfortunately, conflicting judgments on the test to be employed in this context exist.⁶³ In attempting to ‘harmonise’ these disparate views, the ECtHR recently placed unity of facts at the centre of the inquiry, holding that art 4 ‘must be understood as prohibiting the prosecution or trial of a second “offence” in so far as it arises from identical facts or facts which are substantially the same’.⁶⁴ Consequently, the inquiry should focus on those facts which ‘constitute a set of concrete factual circumstances involving the same defendant and inextricably linked together in time and space’, the existence of which must be proved to secure a conviction or institute criminal proceedings.⁶⁵ Double prosecution as well as double punishment is prohibited; therefore one cannot respect the principle by simply reducing the amount of the second punishment by that of the first.⁶⁶

58 Protocol No 7 to the Convention for the Protection of Human Rights and Fundamental Freedoms, as amended by Protocol No 11 (Strasbourg 22 November 1984). Only four member states of the Council of Europe (including the UK and Germany) have not ratified this protocol.

59 On this see Protocol No 7 to the Convention for the Protection of Human Rights and Fundamental Freedoms: Explanatory Report, <<http://conventions.coe.int/Treaty/en/Reports/Html/117.htm>> [32] accessed 12 June 2013.

60 See *Sergey Zolotukhin v Russia*, Application No 14939/03 (ECtHR, 10 February 2009) [52].

61 *A Menarini Diagnostics SRL v Italy*, Application No 43509/08 (ECtHR, 27 September 2011).

62 *Franz Fischer v Austria*, Application No 37950/97 (ECtHR, 29 May 2001) [25].

63 *Gradinger v Austria*, Series A No 328 C (ECtHR, 23 October 1995) [55]; *Oliveira v Switzerland*, Reports 1998 V (ECtHR, 30 July 1998) [25]–[29]; and *Franz Fischer v Austria*, Application No 37950/97 (ECtHR, 29 May 2001) [25].

64 *Sergey Zolotukhin v Russia*, Application No 14939/03 (ECtHR, 10 February 2009) [82].

65 *Sergey Zolotukhin v Russia*, Application No 14939/03 (ECtHR, 10 February 2009) [84].

66 *Franz Fischer v Austria*, Application No 37950/97 (ECtHR, 29 May 2001) [30].

The scope of the principle of *ne bis in idem*, as enshrined in art 4 of Protocol No 7, remains relatively restricted: the principle only applies to ‘acts within the same jurisdiction and does not prevent a person from being convicted for the same offence in a different jurisdiction’.⁶⁷ Parallel antitrust prosecutions by the Commission and an NCA, or by the NCAs of several member states, would therefore fall outside the scope of this provision. Article 4 is therefore only relevant to the first of the manifestations of the additional dynamic contention. Admittedly, however, provided that they are implementing EU law, such parallel prosecutions would not fall outside the scope of *ne bis in idem* articulated in the EU Charter of Fundamental Rights (CFR).⁶⁸ Article 50 CFR therefore extends the relevance of *ne bis in idem* to all three of the valid manifestations of the additional ‘dynamic’ contention.

(ii) European antitrust criminalisation and the relevance of *ne bis in idem*

Ne bis in idem becomes relevant when administrative antitrust sanctions exist alongside criminal sanctions as a broad interpretation has been given to the concept of ‘undertaking’. An undertaking is defined as ‘every entity engaged in an economic activity, regardless of the legal status of the entity and the way in which it is financed’.⁶⁹ The concept includes entities such as trade associations, agricultural cooperatives, sporting bodies, companies and partnerships. *Individuals* can be undertakings if they carry out an economic activity.⁷⁰ There are two exceptions to this however: employees⁷¹ and final purchasers.⁷² If an infringement of art 101 TFEU is found, fines are imposed upon ‘persons’; this is due to art 299 TFEU: ‘[d]ecisions of the Council or of the Commission which impose a pecuniary obligation *on persons* other than States shall be enforceable’.⁷³ Although an undertaking is not always equivalent to a company with legal personality, ‘it is necessary for the purposes of applying and enforcing decisions to identify an entity possessing legal personality to be the addressee of the measure’.⁷⁴ Fines are therefore imposed upon those legal persons to whom the conduct of the undertaking can be imputed. Because art 299 TFEU only refers to ‘persons’, fines can presumably be enforced against both ‘legal persons’ and ‘natural persons’ (that is, individuals).⁷⁵ So, due to both (a) EU jurisprudence on the concept of an undertaking and (b) art 299 TFEU, ‘an individually owned firm and a partnership without a distinct legal personality are undertakings in the sense of EU competition law, and may be fined as such’.⁷⁶

With cartel criminalisation, for a given (alleged) cartel, a natural person in her capacity as an undertaking may be subject to an administrative fine *in addition to* a criminal fine that could be imposed upon her solely as a natural person: unity with regard to the identity of the defendant may arise when both (corporate) administrative sanctions and (personal) criminal sanctions are available. This is where the principle of *ne bis in idem* becomes relevant. For the principle to have effect in practice, two additional elements are required: (i) the existence of (a particular set of) parallel prosecutions; and (ii) the fact that the second

67 Blanco (n 38) [4.44].

68 See art 50 CFR.

69 *Höfner and Elser v Macrotron* (C-41/90) [1991] ECR I-1979 [21].

70 See eg IV/29.559, *RAI/UNITEL* [1978] OJ L157/39 (opera singers) (Commission Decision, 15 June 1978); and IV/28.996, *Reuter/BASF* [1976] OJ L254/40 (individual inventors) (Commission Decision, 26 July 1976).

71 *Albany International BV v Stichting Bedrijfspensioenfonds Textielindustrie* (C-67/96) [1999] ECR I-5751.

72 *FENIN v Commission* (T-319/99) [2003] ECR II-357 [36].

73 Emphasis added.

74 *Limburgse Vinyl Maatschappij and Others v Commission* (T-305–307/94 etc) [1999] ECR II-945, [978].

75 See W Wils, *Optimal Enforcement of EC Antitrust Law: A Study in Law and Economics* (Kluwer Law International 2002) 176.

76 Zuleeg (n 8).

prosecution concerns an offence for which an individual has already been finally acquitted or convicted in the first prosecution (which is termed here 'equivalence of offence').

The existence of parallel prosecutions is the least problematic element. According to art 4(1) of Protocol No 7, the principle may only be relevant when the administrative fine is imposed by the *same jurisdiction* as that imposing the criminal sanction. However, with art 50 CFR, the principle is extended to include antitrust proceedings in other member states. When applicable, art 50 CFR captures two further scenarios: parallel antitrust prosecutions by the Commission and an NCA, and parallel antitrust prosecutions by the NCAs of two or more member states. All three of the valid manifestations are therefore relevant here. The principle will not be infringed where there are parallel antitrust prosecutions involving a member state and a non-EU jurisdiction.

The second element, equivalence of offence, is more complex. Recent jurisprudence requires a factual analysis to be undertaken in situations where the issue of *ne bis in idem* may arise. One must determine the facts required to secure a conviction or institute criminal proceedings in relation to both the administrative and the criminal antitrust offence. If these facts are identical or substantially similar then *ne bis in idem* will apply. Where the (national) criminal antitrust offence merely provides criminal sanctions for violation of art 101 TFEU, there is no real problem: the only difference between the administrative and the criminal proceedings will be the nature of the sanction imposed and not the underlying conduct which is punished; and if the underlying conduct is the same, the facts which are required to convict will also be equivalent. A different outcome may occur, however, when antitrust criminalisation involves the creation of a separate stand-alone criminal offence with elements additional to those included within art 101 TFEU.⁷⁷ The test for equivalence of offence no longer focuses on the essential (definitional) elements of the offences; therefore, the existence of such additional elements does not *ipso facto* ensure the irrelevance of *ne bis in idem* concerns. If, however, the additional elements ensure that a different set of concrete factual circumstances surrounding the cartel are relevant to the criminal antitrust offence from those relevant to the administrative offence, there would be no potential *ne bis in idem* problem. This is a factual inquiry. To determine if the principle is operative, one must: (a) study the facts relied upon in the previous decision (or judgment); (b) examine the facts as stated in the relevant indictment (or the Statement of Objections); and (c) decide whether the facts in (a) and (b) are identical or substantially the same. Where they are not, there is no issue with *ne bis in idem*. When (substantial) factual identity exists, the principle will apply.

(iii) Dealing with *ne bis in idem* in the context of European antitrust criminalisation

The easiest case to deal with occurs where the second antitrust prosecution has begun or is about to begin. In this scenario, when the principle is operative, the charge against the individual (or, alternatively, the undertaking represented by an individual) should be dropped. The reason for this is clear: *double* prosecution is prohibited by art 4(1), and one cannot stay within the principle simply by reducing the amount of the second sanction by that of the first. The situation becomes more complicated when neither of the two parallel prosecutions has begun and a decision is required concerning whether to proceed with a prosecution and, if so, which one, administrative or criminal. (In such a situation, it may also be difficult to establish whether a *ne bis in idem* problem is likely: in practice, one will be precluded from following the three-step methodology noted above, and one may therefore have difficulty in determining the existence of factual identity.) If the prospective criminal and administrative prosecutions will occur in the same member state, then it will be the authorities of that member state which will have to weigh up the relative merits and

⁷⁷ An example is the UK Cartel Offence: s 188 of the Enterprise Act 2002.

demerits of the prospective prosecutions. They should consider resource constraints, likelihood of success, impact upon compliance and, importantly, the degree to which the objectives of the law will be achieved.

When another member state or the Commission is involved (that is, when art 50 CFR becomes relevant), it will be necessary for the authorities to communicate with one another and to resolve the issue through dialogue. Identifying and neutralising any potential *ne bis in idem* problem at the outset is advisable: it avoids the waste of time and resources engendered when cases later collapse on human rights grounds. The ECN may be useful here. Under reg 1/2003, the NCAs, when acting under art 101 TFEU, are obliged to inform the Commission (a) after commencing the first formal investigative measure⁷⁸ and (b) no later than 30 days before taking a decision.⁷⁹ The other member states' NCAs 'may' also be so informed.⁸⁰ Such a flow of information is helpful as it notifies the competition authorities of possible parallel investigations/prosecutions, thereby enabling them to allocate cases amongst themselves to eliminate potential *ne bis in idem* problems. Nonetheless, there are difficulties with the employment of the ECN as a preventive mechanism.

The first difficulty is that there is no obligation concerning inter-NCA notifications. The second is that, due to the vagueness of recital 8 of reg 1/2003, NCAs may not feel obliged to inform the Commission of criminal antitrust cases which involve the imposition of sanctions upon natural persons. Both deficiencies have potential to undermine the ability of the ECN to operate as a preventative mechanism: the authorities of the relevant jurisdictions may not be aware of the potential parallel investigation and prosecution and, hence, the need to discuss appropriate case allocation. Consider the following example. NCA 1 informs the Commission of its (administrative) investigation of undertaking X (an individual); NCA 1 does not inform any other NCA about the investigation; NCA 2 does not inform the Commission (or any NCA) about its prospective criminal case against the individual that constitutes undertaking X. In this scenario, the *ne bis in idem* problem will only become apparent (if detected) once the second prosecution begins; in other words, the ECN would not have been able to operate effectively and eliminate the *ne bis in idem* concern before it developed. Finally, the NCAs cannot relieve each other of the competence to pursue a given antitrust investigation or prosecution. Any resolution of the issue will therefore require the consent of all relevant NCAs.⁸¹ The Commission itself may use art 11(6) of reg 1/2003 to investigate the (individual comprising the) undertaking itself, thereby relieving an NCA of its competence to apply art 101 TFEU. While this power may appear useful in preventing the emergence of *ne bis in idem* problems, it too suffers from limitations. Importantly, art 11(6) is likely only to be used in very exceptional circumstances and only where the EU interest so dictates: it is not a panacea when conflict between NCAs develops.⁸² Furthermore, art 11(6) only relieves the NCAs of their competence to apply art 101 TFEU; it does not relieve criminal courts of their power to impose sanctions for violation of the criminal law when such courts are not NCAs for the purpose of reg 1/2003. Moreover, if a national cartel offence is not actually a 'national competition law' for the purposes of reg 1/2003,⁸³ then art 11(6) will not have any impact on the national criminal proceeding as reg 1/2003 will not apply. It is not inconceivable, then, that potential

78 Reg 1/2003, art 11(3).

79 Ibid art 11(4).

80 Ibid arts 11(3) and 11(4).

81 On the ability of the NCAs to terminate their own antitrust proceedings when there is a parallel (art 101 TFEU) proceeding in another member state, see reg 1/2003, art 13.

82 See generally European Commission (n 30) s 3.2.

83 On this, see *IB v The Queen* [2009] EWCA Crim 2575.

ne bis in idem problems will not be neutralised before the commencement of the parallel prosecutions. This effectively returns one to the first situation examined above: if parallel antitrust prosecutions are not prevented by the competition authorities, then the charges in the second prosecution should be dropped to respect due process.

(iv) Consequences of respecting the principle of *ne bis in idem*

If cartel activity is criminalised for retributive reasons, one would have to establish measures to ensure that a cartelist is protected from excessive punishment for her wrongful conduct: the theory of retribution ensures that a person shall be subjected to no more punishment than she deserves. One such measure could be a judicially recognised prohibition on an individual being tried/punished more than once for the same wrongful conduct. The principle of *ne bis in idem* can therefore be rationalised in part using retribution theory. As noted above, however, the principle can require one to drop the second antitrust prosecution to respect due process. This fact influences the effectiveness of an enforcement policy premised on the achievement of deterrence. The reason for this is as follows. If the enforcement action represented by the first prosecution is indeed an administrative proceeding, it will be the criminal proceeding that will have to be dropped. The administrative proceeding, then, will be the only proceeding against the individual cartelist (who is also an undertaking). Significant deficiencies concerning the use of administrative antitrust sanctions alone to secure deterrence exist, including both the inability of the competition authority to impose an optimal fine and the inability of the undertaking to pay such a fine.⁸⁴ As, under the circumstances identified, only an administrative proceeding is legally possible, these deficiencies would continue to be relevant. Consequently, sub-optimal deterrence may result when, due to *ne bis in idem*, only an administrative sanction is imposed for a given cartel.

(C) CONCURRENT PROCEEDINGS

All three of the valid manifestations of the additional dynamic contention are relevant to concurrent proceedings: they all concern two separate antitrust proceedings which take place either simultaneously or within a short period of time of each other. Such concurrent proceedings raise two due process-related issues.

(i) Two due process-related issues

The first due process-related issue arises because the substance of due process rights (at least under the ECHR) is not exactly equivalent as regards both criminal and administrative regimes.⁸⁵ Assume that an antitrust authority is conducting an administrative investigation against an undertaking using its administrative powers. During that administrative proceeding, the antitrust authority uncovers evidence implicating a particular individual cartelist. The antitrust authority then decides to commence a criminal proceeding against that individual. If the due process rights protected under the administrative proceeding are not as robust as those to be respected under a criminal proceeding, the evidence collected using the administrative powers could not be used in a criminal court without violating art 6 ECHR. Of course, if evidence is collected to a criminal standard and is then used in an administrative proceeding, there should be no human rights-based objection: the strictest of legal guarantees will have been observed and potential prejudice will be absent.⁸⁶

84 See Whelan (n 23).

85 On these differences, see Whelan (n 6).

86 See eg HL Deb 29 July 2002, vol 638, col 73 (Lord Sainsbury).

The second due process-related issue concerns the creation of prejudicial pre-trial publicity. Such publicity may be created if an administrative cartel proceeding *precedes* any related criminal proceedings against an individual. The reason for this should be clear: if a (public) administrative decision determines that a market has been cartelised, it may be publicised by the international media, particularly if the market is important for consumers. The media could be critical of not only the undertakings involved, but also of their staff. Consequently, such negative publicity may influence potential jurors in a later criminal trial involving an individual employee of that undertaking. Such publicity could engender impartiality in the court to the detriment of compliance with art 6 ECHR.⁸⁷

(ii) Ensuring consistency with due process

To avoid the first due process-related issue, one should not present evidence in a criminal court if it is gathered only to the administrative standard, otherwise one will likely violate art 6 ECHR. Criminal antitrust proceedings should respect the rights of the accused to a criminal standard *at all stages* of the investigation and prosecution. This is not to say, however, that parallel (criminal and administrative) investigations by the same public antitrust body in a given member state cannot be conducted.⁸⁸ In fact, such parallel enforcement actions can be successfully conducted if mechanisms exist to avoid the relevant due process-related issue.

The first required mechanism is a system which ensures from the outset that the administrative investigative team is functionally and materially separate from the criminal investigative team. Under this system, each investigative team acquires evidence according to its powers and respects the appropriate level of due process rights (administrative or criminal). What is to be avoided here is the ‘fusion’ of the investigations (ie, the treatment of the administrative and criminal investigations as one single investigation), not their parallel advancement (by separate teams within the antitrust enforcement agency).⁸⁹

A second mechanism should exist to ensure that this division of enforcement teams remains solid throughout both investigations. If different teams within the one authority are pursuing respectively administrative and criminal investigations, ‘Chinese walls’ must be established between these teams.⁹⁰ In particular, if information or evidence is to move from one investigative team to another, then it should first pass through a screening process involving lawyers who are not part of either investigative team. These lawyers should be qualified to appreciate whether, in passing on the relevant information or evidence and allowing it to be used by others, due process will be undermined. Where this is likely, the information or evidence in question should not be passed on. Legislation (and/or soft law)⁹¹ could also be adopted to provide clarity to antitrust enforcers on legitimate information exchange. Such legislation could detail, for example, the circumstances in which information passed from the administrative team to the criminal team can be legitimately relied upon in court.⁹²

87 For ECtHR jurisprudence on impartiality contrary to art 6 ECHR, see eg *De Cubber v Belgium* (1984) 7 EHRR 236 [24]; *Gregory v UK* (1997) 25 EHRR 577 [45]; and *Hauschildt v Denmark* (1989) 12 EHRR 266 [46].

88 See Nazzini (n 9).

89 Ibid 485.

90 M Frese, ‘The Negative Interplay Between National Custodial Sanctions and Leniency’ in K Cseres, P M Schinkel and F Vogelelaar (eds), *Criminalization of Competition Law Enforcement: Economic and Legal Implications for the EU Member States* (Edward Elgar Publishing 2006) 206. On the (successful) use of Chinese walls in the UK criminal cartel regime, see E Morgan, ‘Criminal Cartel Sanctions under the UK Enterprise Act: An Assessment’ (2010) 17(1) *International Journal of the Economics of Business* 67, 72.

91 See eg OFT, *Powers for Investigating Criminal Cartels* (OFT 515, January 2004) [4].

92 See eg Competition Act 1998 (CA), s 30A

Finally, to ensure efficiency in enforcement, a formalised decision-making process should be initiated which will determine whether an administrative or criminal proceeding (or both) should be followed for a given cartel. If it is unclear initially which type of proceeding should be initiated, evidence should be collected to the criminal standard. Such an approach exists in the UK: the Office of Fair Trading (OFT), when pursuing cartel cases, often ensures that criminal standards are followed, even in the course of an administrative proceeding.⁹³ It is only when it becomes clear that an administrative proceeding should also be initiated (or should take the place of the criminal investigation) that the administrative standards should be deemed relevant. By creating a formal structure (composed of experienced investigators and/or prosecutors) which determines whether an investigation should be criminal or administrative (or both), one can ensure that due process is not violated, while limiting the amount of resources spent on unnecessary investigative efforts. It can also help to avoid situations where evidence is initially collected using administrative powers but which then must be reacquired using criminal powers.

To avoid the second due process-related issue, one could ensure that the outcome of the criminal proceeding is resolved before that in the administrative proceeding; negative publicity is less likely to affect a criminal trial in the absence of a prior formal public decision concerning the actual existence of the cartel in question.⁹⁴ Admittedly, such an approach may not always be possible. For example, the administrative proceeding may take place in a different jurisdiction than the criminal proceeding.⁹⁵ In such a situation, it would be open to the criminal jurisdiction to request that the administrative proceeding in the other jurisdiction be stayed pending the outcome of the criminal trial; but such a request would not have to be granted by the other jurisdiction (or may not be possible under its law).

But this may not necessarily be a problem. As (negative) publicity surrounding cartel decisions is often of limited duration and of limited interest to the public, one cannot assume that it automatically becomes prejudicial to a criminal trial. Second, the antitrust authority could simply de-emphasise the importance of the administrative proceeding (for instance, by not commenting on it publicly) in an attempt to limit the interest of the national media. This option is particularly relevant when only the first manifestation of the additional dynamic contention is operative: the antitrust authority will have more influence over the extent of publicity surrounding the administrative antitrust decision than it would have if it was adopted by a foreign authority (which presumably would have a strong interest in publicising its enforcement results). Third, to reduce the negative potential of publicity, one could delay the criminal cartel trial for a short period of time to allow a 'fade effect' to develop.⁹⁶ Too long of a delay may, of course, engender problems. For example, the memories of witnesses may also begin to fade with time. Finally, even if one could establish that publicity may influence decision makers in the criminal cartel trial, it would still be difficult to establish a legal challenge to that trial alleging a violation of art 6 ECHR. The reason for this is that under ECtHR jurisprudence, the existence of negative publicity per se is not decisive concerning impartiality; one should also consider the effectiveness of measures offered to exclude any legitimate doubt concerning prejudice.⁹⁷ As noted by Lord Hope:

93 See OFT (n 91) [4.1] and [4.9]; and OFT, *The Proposed Criminalisation of Cartels in the UK: A Report Prepared for the Office of Fair Trading by Sir Anthony Hammond KCB QC and Roy Penrose OBE QPM* (OFT 365, November 2001) [1.20].

94 See OFT, *Proposed Criminalisation* (n 93) [1.20].

95 Ibid.

96 See Nazzini (n 9) 486.

97 *De Cubber v Belgium* (1984) 7 EHRR 236 [24]. See also *Pullar v UK* (1996) 22 EHRR 391 [40].

the question is not confined to the residual effect of the publicity on the minds of each of the jurors. Account must also be taken of the part which the judge will play to ensure, so far as possible, that the defendants will receive a fair trial. An examination of the measures which [she] can take under the system which has been laid down for the conduct of criminal jury trials forms an integral part of the whole exercise.⁹⁸

The relevant measures include, in particular, warnings or directions provided to the jury by the trial judge.⁹⁹ For an argument concerning prejudicial pre-trial publicity to be successful in a criminal cartel case, then, one must convince a court that the available judicial measures are ineffective in protecting against impartiality. This may be difficult to do in practice, even if academic literature supports the argument:¹⁰⁰ the effectiveness of judicial directions may represent a ‘convenient fiction’¹⁰¹ which the judiciary is generally content to accept.

(iii) Impact upon the objectives of criminalisation

The main reason that the techniques identified are capable of affecting deterrence is that to implement these techniques one must expend resources. Additional resources are inevitably consumed when: (a) evidence collected to an administrative standard is then recollected to a criminal standard; (b) the investigation respects a (more demanding) criminal standard from the outset instead of an administrative standard (even when, ultimately, an administrative sanction results); and (c) mechanisms, such as Chinese walls, are formalised, as staff independent of the investigative teams are required. Such increases in costs are relevant to economic deterrence as such an objective seeks to achieve *economic efficiency*. These additional costs need to be fed into the cost–benefit analysis conducted when deciding whether to introduce criminal antitrust sanctions and further complicate the decision whether to criminalise. The real question that arises is whether the additional costs concerning the techniques can be offset by the increase in benefits (due to marginal deterrence) secured by the criminalisation. This is a factual issue. The point here is that the protection of due process does in fact influence that particular analysis; hence such protection affects the objective of deterrence.

The techniques which may impact upon the achievement of retribution are those related to the protection against prejudicial pre-trial publicity. When pursuing the objective of deterrence alone, one does not necessarily have to protect against negative publicity: (general) deterrence theory per se does not require that only the guilty are punished, unless such (undeserved) punishment is capable of undermining the general level of respect for the criminal law and therefore its deterrent effect. In fact, if citizens genuinely believe that an individual is guilty of a criminal cartel offence, her conviction (even if erroneous) may increase their respect for the law and may increase general deterrence. With retribution, by contrast, such an approach is prohibited: only *deserved* punishment can be imposed; negative pre-trial publicity which leads to an *erroneous* conviction is inconsistent with this objective. To the extent, then, that the identified techniques concerning the avoidance of prejudicial pre-trial publicity are effective, such techniques can ensure consistency with retribution.

98 *Montgomery v HM Advocate* [2003] 1 AC 641, 670.

99 See T M Honess, ‘Empirical and Legal Perspectives on the Impact of Pre-Trial Publicity’ [2002] *Criminal Law Review* 719, 720.

100 See eg P Fitzpatrick, ‘The British Jury: An Argument for the Reconstruction of the Little Parliament’ [2010] *Cambridge Student Law Review* 1. See also C Thomas, ‘Are Juries Fair?’ (Ministry of Justice Research Series 1/10, February 2010) <www.justice.gov.uk/publications/are-juries-fair.htm> accessed 12 June 2013, demonstrating that a high proportion of jurors do not understand judicial directions fully in the legal terms used by judges.

101 J A Tanford, ‘The Law and Psychology of Jury Instructions’ (1990) 69 *Nebraska Law Review* 71, 111.

Nonetheless, one technique could undermine retribution: limiting the interest of the national media in the concurrent administrative antitrust proceeding. This technique is problematic as, given the nature of cartel activity, 'norm-building' will be required during the early stages of criminal enforcement to operationalise the objective of retribution.¹⁰² In particular, the authorities must create awareness of the moral content of cartel activity. The easiest way to do this is to adopt administrative decisions condemning cartel activity while creating publicity surrounding such decisions. When such publicity is shunned, efforts to create awareness of the moral content of cartel activity will be undermined. This problem can be avoided, however, if the administrative antitrust proceeding follows the criminal one. Accordingly, maximum publicity can be created concerning the administrative proceeding without damaging the sustainability of the criminal conviction. It is only when the administrative antitrust proceeding actually precedes the criminal proceeding that the tension between the objective of retribution and respect for due process is created.

Conclusion

Due process presents difficult challenges for those considering the introduction of criminal antitrust sanctions within the EU. Some of these challenges are the natural result of the additional dynamic contention, which holds that the introduction of criminal antitrust sanctions in a European jurisdiction does not preclude the additional imposition of administrative sanctions for a given cartel. The validity and impact of the additional dynamic contention were examined to present a more comprehensive understanding of the legal challenge of due process for European antitrust criminalisation.

The validity of the additional dynamic contention was considered first. Three different sentencing scenarios for a given cartel were found to be legally possible and likely in practice. The additional dynamic contention was therefore found to be valid. The next analytical step determined the actual impact of the additional dynamic contention and, hence, the extent of the challenge presented by due process. Three issues were considered: the exchange of information; double jeopardy; and concurrent proceedings.

A potential due process issue was identified with the exchange of information from administrative antitrust regimes to criminal antitrust regimes: if such information could be used in evidence in a criminal trial, it could potentially undermine due process, as the protections in an administrative regime may not be as robust as those in a criminal regime. While EU law has foreseen this problem, its response arguably goes too far: it prohibits the use of information in evidence to secure custodial sanctions in a receiving (criminal) regime even where the sending (administrative) regime has respected the human rights guarantees in the receiving state. To ensure that the objectives of antitrust criminalisation are not undermined, this custodial rule should be abolished. Alternative, less strict protections could also be introduced, such as the use of a double barrier. However, if due process is indeed to be protected, it is necessary to accept an inevitable negative impact upon the efficiency of the criminal antitrust regime: legal appeals questioning whether the relevant collection of information respects the equivalent level of protection of the rights of the accused in the receiving jurisdiction inevitably consume resources. Compensating for this reduction in efficiency represents an inherent legal challenge of European antitrust criminalisation.

The second identified issue of due process concerned double jeopardy. This issue becomes relevant due to: (a) the validity of the additional dynamic contention; and (b) the fact

¹⁰² See P Whelan, 'Morality and its Restraining Influence on European Antitrust Criminalisation' (2009) 12 *Trinity College Law Review* 40.

that an individual can constitute an undertaking under EU law. One must avoid those rare cases where double jeopardy issues arise when both administrative and criminal sanctions are imposed for a given cartel. The least difficult situation occurs when the second prosecution has not yet begun, but the first prosecution has: the second prosecution should be dropped. When neither prosecution has begun, the situation becomes problematic, particularly where more than one jurisdiction is involved. The ECN can be used to resolve the issue. But three limitations were identified with this solution: (i) there is no obligation concerning inter-NCA notifications within the ECN; (ii) the NCAs may not feel obliged to inform the Commission of antitrust cases involving the imposition of sanctions upon individuals; and (iii) the NCAs cannot relieve each other of the competence to pursue antitrust cases. These limitations, then, also contribute to the inherent legal problem of due process.

The final identified issue of due process involved concurrent proceedings. Two potentially problematic matters were identified: (a) the use of administrative powers to collect evidence for a parallel criminal investigation; and (b) the engendering of prejudicial pre-trial publicity. Techniques can be used to reduce the likelihood of their occurrence. However, efforts to ensure consistency with due process may impact negatively upon the objectives of European antitrust criminalisation. Specifically, such techniques may either increase the costs involved in criminal antitrust enforcement, thereby reducing the possible net (deterrence-based) gain from criminalisation, or require the authorities to limit media interest in the (concurrent) administrative proceeding, to the detriment of retribution. The limitations of the identified techniques further contribute to the inherent legal problem of due process.

The analyses which are presented in this article should help to inform the current debate on European antitrust criminalisation. It is clearly important in the context of this debate that one can establish that neither EU law itself nor the practical challenge of adhering to the dictates of the ECHR precludes in any absolute sense the imposition of criminal cartel sanctions alongside their administrative counterparts for a given cartel. It is significant that procedures can be put in place to minimise the concerns that art 6 ECHR will be violated where one jurisdiction imposes criminal cartel sanctions and another imposes administrative cartel sanctions for a given cartel. This fact ensures that some European jurisdictions can forge ahead with antitrust criminalisation even when a consensus does not exist across Europe concerning the necessity and appropriateness of such criminalisation: it facilitates a slow, measured conversion to antitrust criminalisation for those European jurisdictions which are materially ready for the introduction of criminal cartel sanctions.

It is true too that the option of having both criminal and administrative sanctions on their legislative books is arguably an attractive option for some European jurisdictions with a relatively developed competition law culture. The current literature establishes that robust arguments can in fact be put forward as to why criminal sanctions on individuals are superior to administrative sanctions on undertakings in achieving deterrence of cartel activity. However, there are also good deterrence-focused reasons for reserving the option of imposing administrative sanctions on undertakings in addition to criminal sanctions on individuals for a given cartel: the absence of a sanction on the company would incentivise the company to encourage cartel activity by its employees as such activity would ultimately benefit the undertaking concerned. Consequently, those European jurisdictions which are serious about enforcing their cartel laws may take comfort in the fact that the option of imposing administrative sanctions for a given cartel can still remain a legal possibility even when criminal sanctions have been deployed against individuals involved in that cartel.

None of this is to say, however, that the additional dynamic does not present challenges concerning adherence to art 6 ECHR when enforcing competition law. Indeed, as

articulated by the analyses presented in this article, the requirement to adhere to art 6 ECHR is capable, in certain circumstances, of impacting upon both objectives of European antitrust criminalisation: deterrence and retribution. By explaining how and when this is so, this article makes a further, novel contribution to the debate on European antitrust criminalisation. Without doubt, those contemplating the introduction of criminal cartel sanctions should be aware of the inherent challenges of adhering to the relevant due process requirements. As noted above, one aspect of due process has been considered in detail in the literature: the claim that the use of criminal cartel sanctions in the place of administrative cartel sanctions inevitably results in a 'strengthening of rights' to the benefit of the accused and to the detriment of the relevant enforcement objectives. However, to understand fully the challenge of due process for European antitrust criminalisation, one must understand whether the imposition of criminal cartel sanctions alongside administrative cartel sanctions for a given cartel presents its own due process challenges. By articulating the specifics of this particular aspect of the challenge of due process, this article allows for a more informed view as to the legal challenges represented by antitrust criminalisation and therefore should help European jurisdictions to come to their own conclusion whether such criminalisation is in fact worth the effort that needs to be expended to make it legally sound in practice.

Carbon sequestration and the limits of regulation: a UK perspective

MARK STALLWORTHY

Swansea University

1 Introduction

Considerable political capital is invested in developed economies in the idea of carbon capture and storage (CCS), with a heavy weight of expectation upon eventual deployment.¹ The UK government describes it as ‘key if the UK is to achieve our climate change targets whilst maintaining a vibrant, competitive economy’.² The purpose behind CCS is to mitigate impacts from burning fossil fuel energy sources by diverting carbon emissions away from atmospheric release. Complex technology applies, but simply put, CCS involves capture (by chemical and other means) of carbon dioxide (CO₂) from major point source installations,³ to be compressed and transported, generally via pipelines, to designated sites (principally in depleted oil and gas fields, or saline aquifers), for permanent storage by injection into deep geological formations, below ground or under the seabed.⁴ Thus is held out the enticing promise of at last a significant technological response to climate change; one which yet remains unproven as a composite process to a large scale. Yet if not as controversial as are certain other proposed geo-engineering solutions,⁵ CCS may prove a poisoned chalice, insofar as political expectations for it can slow development of alternative energy strategies. By contrast with more developed areas of climate change law, and arguably reflecting the still inchoate existence of the physical process itself, a body of

1 See eg IEA, *Carbon Capture and Storage: Model Regulatory Framework* (OECD/IEA 2010) 1.1.1.

2 Department of Energy and Climate Change (DECC), *CCS Roadmap: Supporting Deployment of Carbon Capture and Storage in the UK* (April 2012) 2.5 <www.gov.uk/government/uploads/system/uploads/attachment_data/file/48317/4899-the-ccs-roadmap.pdf>.

3 Mainly power generation, but also other energy-intensive processes, such as steel and cement production. See examples of demonstration projects <www.globalccsinstitute.com/projects/browse>.

4 Saline aquifers seem to offer the greatest opportunities within Europe: it is estimated as sufficient for around 100 years’ worth of Europe’s emissions. A long-running pilot pre-combustion project (since 1996) has operated at the Sleipner natural gas processing facility in Norway, with around 1 million tonnes of CO₂ extracted from natural gas and pumped into saline aquifers: see <<http://statoil.com/en/>>.

5 Cf J Goodell, *How to Cool the Planet: Climate Engineering and the Audacious Quest to Fix the Earth’s Climate* (Houghton Mifflin Harcourt 2010).

legal scrutiny has been slow to emerge,⁶ although in the wake of legislative efforts within Europe, we are now seeing a more focused attention to regulatory implications.⁷

The potential value of CCS is evident from reports that global emissions levels remain on a stubbornly upward trajectory,⁸ with rising demand for fossil fuels,⁹ especially for power generation.¹⁰ Global trends encapsulate a free rider problem that undermines any achievements that result in those jurisdictions that have in place emissions reductions policies supported by legal commitments.¹¹ That said, experience in the UK suggests that claims for successful achievement of reductions require careful scrutiny.¹² Further, of the UK power sector's proportion of 27 per cent of total UK emissions,¹³ around four-fifths of that figure remains reliant on fossil fuels.¹⁴ Meanwhile, globally, research carried out by the International Energy Agency (IEA) suggests that, with electricity-related CO₂ emissions expected to double between 2007 and 2050, then to meet international objectives for a 50 per cent reduction in aggregate emissions to 2050, CCS would be required to contribute as much as one-fifth of that total reduction.¹⁵ CCS thus presents, if not quite a carbon version of the last chance saloon, then a potentially pivotal technological contribution toward climate change mitigation.¹⁶ Further, if it proves possible to establish CCS projects

-
- 6 Although much attention is evident in relation particularly to property law implications from US perspectives; this can be said to reflect such locational features as an abundance of sub-terrestrial sites with storage potential, alongside experience of smaller-scale technologies such as enhanced oil recovery, plus a high incidence of private rather than state ownership of underground rights: see A B Klass and E J Wilson, 'Climate Change, Sequestration, and Property Rights' [2010] *University of Illinois Law Review* 363; N R Hoffman, 'The Feasibility of Applying Strict-Liability Principles to Carbon Capture and Storage' (2010) 49 *Washington Law Journal* 527.
 - 7 I Havercroft, R Macrory and R B Stewart (eds), *Carbon Capture and Storage: Emerging Legal and Regulatory Issues* (Hart Publishing 2011) offers an excellent introduction to key issues and comparative experiences to date; see also A Haan-Kamminga, M A Roggenkamp, E Woerdman, 'Legal Uncertainties of Climate Capture and Storage in the EU: The Netherlands as an Example' (2010) 4(3) *Carbon and Climate Law Review* 240; A Haan-Kamminga, 'Long Term Liability for Geological Carbon Storage in the European Union' (2011) 29(3) *Journal of Energy and Natural Resources Law* 309; G Kaminskaite Salters, 'Implementing the Carbon Capture and Storage Regime in the UK: The Challenges and Solutions' in M Peeters, M Stallworthy and J de Cendra de Larragan (eds), *Climate Law in EU Member States: Towards National Legislation for Climate Protection* (Edward Elgar Publishing 2012) ch 12.
 - 8 For a survey of the scale of the problem and proposals for transforming energy sourcing and use, see T B Johansson et al (eds), *Global Energy Assessment* (International Institute for Applied Systems Analysis Report, CUP 2012).
 - 9 Especially important where significant commitments to coal-fired generation persist: see eg Office of Management and Budget, *A New Era of Responsibility: Renewing America's Promise* (26 February 2009) <www.gpoaccess.gov/usbudget/fy10/pdf/fy10-newera.pdf>.
 - 10 In some states there is very high dependence on coal: eg over 90% in South Africa and Poland; near 80% in China, Australia: see <www.worldcoal.org/resources/coal-statistics>.
 - 11 For instance, under the UK's Climate Change Act 2008.
 - 12 According to the UK's Climate Change Committee, in 2011 (though the impact of low economic growth, rising fuel prices and a mild winter contributed to a substantial economy-wide fall in emissions) policy measures contributed only a circa 0.8% reduction, against an annual target of 3.0%: *Fourth Annual Progress Report to Parliament* (Climate Change Committee, July 2012) 51.
 - 13 UK emissions statistics 2010 <www.decc.gov.uk/en/content/cms/statistics/climate_stats/gg_emissions/uk_emissions/2010_final.aspx>.
 - 14 Figures for 2010 were approximately as follows: gas (40%), coal (32%), nuclear (18%), renewables (7%), oil (3%): www.decc.gov.uk/en/content/cms/statistics/energy_stats/source/electricity/electricity.aspx.
 - 15 IEA (n 1) 1.1.
 - 16 B Metz et al (eds), 'Summary for Policy Makers' in *Special Report on Carbon Dioxide Capture and Storage* (IPCC Working Group III, CUP 2005) 2–14.

at large scale, then the current terms of debate over energy mix would be radically altered.¹⁷ This is partly on account of the abundance of coal (and to a lesser extent, gas) relative to other fuel sources.¹⁸ It also reflects circumstances in which (for the medium term at least), as compared with fossil fuels, most renewables remain significantly less competitive.¹⁹ Accordingly, a further challenge, aside from aspects of direct regulation, is to secure emissions cuts in spite of prevailing market conditions being favourable to continued exploitation of fossil fuels.

This article addresses key elements of the wider regulatory context into which CCS must be accommodated. The deepening reach of modern environmental regulation, marked by a gradual blurring of distinctions as between the private and the public,²⁰ represents an aspect of what has been termed regulatory capitalism.²¹ An important side effect is that, whilst under pressure in an integrated policy arena, environmental law regimes remain relatively robust, arguably running counter to effects elsewhere from a 'hollowing out of the state, privatisation and regulation'.²² This has been evident in those areas that involve high levels of uncertainty, now most vividly in the sphere of climate change.²³ Risks inherent to CCS include not only significant costs exposures at its various stages of development, operation and regulatory compliance, but also a range of potential liabilities, both established and novel. The complex technologies involved and the potential exposure to liabilities, regulatory and otherwise, present major investment uncertainties. Whilst private investment and expertise appears essential for development and deployment, the state has vital economic and legal roles to play in assuring viability. Though most policy attention has focused on commercial risks in initial development, a state of uncertainty surrounds those contingent risks that attach to eventual storage, notwithstanding a harmonised framework established within the European Union (EU),²⁴ now in course of

17 For instance, already demanding EU renewables targets may require revision in the event of successful technological developments, such as CCS (and in some jurisdictions further nuclear power investment): D Helm, 'EU Climate-Change Policy: A Critique' in D Helm and C Hepburn (eds), *The Economics and Politics of Climate Change* (OUP 2010) 237.

18 A feature of progress in extraction technologies is that figures do not remain constant, but at current rates of consumption, there may be over 120 years of available coal, though some estimates suggest several centuries' worth; cf though with a similar caveat, figures which suggest barely half-a-century for oil and in excess of 60 years or more for gas: see eg <www.worldenergy.org>. The latter figure in particular will require review in light of progress in the development of the technology known as 'hydraulic fracking', already extensive in the US: see R McKie, 'Fracking is Only Way to reach Obama's Goal on Climate Change, Says Top Scientist', *Observer* (London, 17 February 2013). The technology is expected to have a more limited relevance for the UK, although despite concerns at greenhouse damage from methane 'fugitive' emissions and risks of seismic tremor, the UK government has lifted a moratorium following review by the Royal Society/Royal Academy of Engineering: *Shale Gas Extraction in the UK: A Review of Hydraulic Fracturing* (RS/RAE, June 2012).

19 The burgeoning development of fracking in the US can also be expected to have collateral effects, including as lower-cost US coal penetrates world markets to a significant degree.

20 Cf E Fisher, 'Unpacking the Toolbox: Or Why the Public/Private Divide is Important in EC Environmental Law' in M Freedland and J-B Auby (eds), *The Public and Private Law Divide* (Hart Publishing 2006) 230–34.

21 See N Gunningham, 'Environmental Law, Regulation and Governance: Shifting Architectures' (2009) 21 JEL 179, who suggests a resulting 'expansion rather than the retreat of regulation in new forms'.

22 Ibid 209.

23 See R J Lazarus, 'Super Wicked Problems and Climate Change: Restraining the Present to Liberate the Future' (2009) 94 Cornell Law Review 1153 (2009).

24 Dir 2009/31/EC on the geological storage of carbon dioxide [2009] OJ L140/114; affording wide flexibility available to member states in setting financial security requirements for operators, through operational and post-closure (up until transfer of responsibility to the state) phases, to ensure that obligations under permit are met: arts 7(10), 19 (including as to liability arising from leakage in accordance with Dir 2003/87/EC establishing a scheme for greenhouse gas emission allowance trading within the Community [2003] OJ L275/32).

transposition by those member states that regard CCS as both feasible and acceptable.²⁵ There, it is notable that the state should perform the role of ultimate guarantor of long-term liability exposure.²⁶ The approach contrasts with a more general reticence toward state underwriting of liabilities in cognate fields; seen, for instance, in the EU's environmental liability regime,²⁷ and also domestically in governmental resistance to state underwriting of flood losses affecting newly at-risk areas (also linked to effects of climate change).²⁸

Accordingly, as part of the regulatory task, finding an appropriate balancing as between state and private responsibilities will be crucial to securing effective implementation and regulation of CCS into the long term. The article proceeds as follows. Following an introductory review of underlying rationales, the argument will focus on three main themes. Whilst each is introduced separately, across a policy area that (in common with many new generation environmental challenges) fully justifies a polycentric labelling, there is considerable crossover between them.²⁹ They address, first, the wider environmental governance context into which debates over CCS must take place; secondly, instrumental challenges raised by CCS in light of those public and market priorities that influence climate change and energy policy-making; and thirdly, key long-term risk regulation issues, especially concerning potential liabilities, that need to be addressed. Particularly in this latter respect, consideration of the appropriateness of command-and-control and economic regulatory mechanisms is offered, with specific consideration of longer-term risks (from leakage from storage sites) and drawing more general lessons from experience in other fields of environmental regulation.

2 Rationale for CCS: its potential contribution to climate objectives

Aspects of CCS technology are well known in discrete industrial applications, such as in enhanced oil recovery.³⁰ These mostly operate on a small scale, yet to be applied in commercial power generation.³¹ Eventual deployment could take place by one of three identified capture methods.³² The first, a *pre*-combustion process (previously used in large-scale chemical and fertiliser production) could be applied in power generation by producing hydrogen for use in gasification systems.³³ A second process, operating *post*-combustion, follows relatively well-known technology, already in power generation by removal from flue

25 The UK's established regulatory framework will apply to all CCS-related activities, save the permitting of permanent storage: see Storage of Carbon Dioxide (Licensing etc) Regulations 2010/2221.

26 Ibid arts 18 (transfer of obligations from operator to state) and 20 (financial mechanism to cover a limited range of post-closure costs).

27 Dir 2004/35/EC on environmental liability with regard to the prevention and remedying of environmental damage [2004] OJ L143/56: see G Winter, J H Jans, R Macrory and L Kramer, 'Weighing up the EC Environmental Liability Directive' (2008) 20 JEL 163.

28 See J Morgan and M Stallworthy, 'Indemnifying against Flood Loss in a Changing Environment' [2013] 33 Legal Studies 239–63.

29 By contrast with typical point-source environmental risks, climate threats are causatively highly diffuse, with associated harms widely displaced both spatially and temporally: see H Bulkeley and P Newell, *Governing Climate Change* (Routledge 2010).

30 P M Marston and P A Moore, 'From EOR to CCS: The Evolving Legal and Regulatory Framework for Carbon Capture and Storage' (2008) 29 Energy Law Journal 421.

31 For a breakdown of the current status internationally of around 78 large-scale integrated projects, see <www.globalccsinstitute.com/resources/data>.

32 For succinct explanations, see N Shilling, 'Carbon Capture and Storage: An Equipment Manufacturer's Perspective' in Havercroft et al (n 7) 27–30; also K Thambimithu et al, 'Capture of CO₂', in Metz et al (n 16).

33 Fuel reacts with oxygen and air to produce a synthetic gas (syngas), mainly of hydrogen and carbon monoxide (CO), and the latter further reacts with steam to convert CO₂ to be extracted (where coal is the fuel, termed an integrated gasification combined cycle).

gases of pollutants such as sulphur (as well as in small-scale processes, such as soft-drinks carbonisation).³⁴ A third alternative, oxy-fuel (or advanced) combustion, so far restricted to laboratory and small-scale pilots, would occur *during* combustion.³⁵ If successfully demonstrated at large scale, each might be suitable for new installations.³⁶ In contrast, for existing capacity, the pre-combustion method is not compatible with existing solid fuel combustion processes;³⁷ although capture potential exists for post- or oxy-fuel combustion (respectively by retro-fit or re-powering of installations).³⁸

The main justifications for CCS lie in support of (not always compatible) objectives for energy security and emissions reduction. All the above processes offer prospects for emissions reductions: the post- and pre-combustion methods with carbon capture rates of up to 90 per cent, and oxy-fuel combustion with an even greater potential for up to 98 per cent. The future role of fossil fuels could thereby be transformed, somewhat counter-intuitively contributing to low-carbon objectives, as part of the energy mix: with increased availability of 'acceptable' sources, within a more balanced energy portfolio. Further, the use of fossil fuels alongside CCS can ensure (unlike, for instance, wind power, which has the drawback of intermittency) a relatively constant reliability of supply (that is, baseload operation). Less quantifiably, CCS could, if effective, also obviate uncertain risks (and costs) from further diversion of effort into seeking alternative technological breakthroughs.

CCS is therefore heralded as part of a future low-carbon economy, objectives for which underpin the UK's legislative regime for limiting greenhouse emissions.³⁹ Central government must meet mandatory statutory reductions, across periodic carbon budgetary periods. Policy initiatives in accord with this process have resulted in a commitment to rapid decarbonisation of power generation during the 2020s; and the statutory reporting body monitoring progress has set a 2030 objective for the sector of net carbon emissions close to nil.⁴⁰ On current trends, however, this needs to be viewed against a total UK annual generation capacity requirement of over 80 GW.⁴¹ Further, by the early 2020s around a quarter of generation capacity will require replacing⁴² in response to numerous factors: including closure requirements (by 2015) under the Large Combustion Plants Directive,⁴³ stricter sulphur and nitrogen dioxide industrial emission limits,⁴⁴ and anticipated end-of-life

34 Once the fuel has been burned, flue gas is cooled and reacts with a solvent allowing CO₂ to be removed by means of a scrubbing process.

35 Nitrogen is removed from the flue gas when the fuel is burned in 95% pure oxygen, producing a flue gas of water vapour, which is then condensed, alongside CO₂ which can then be separated off.

36 Currently, the most advanced demonstration (coal) plants each aim to divert CO₂ for oil recovery (from 2014): Boundary Dam, Saskatchewan, Canada (post-combustion) and Kemper County, Mississippi, USA (pre-combustion): see eg 'North American Success', *ENDS Report* 445 (London, 22 February 2012) 28.

37 Reliant on burning pulverised coal.

38 Although in the latter case dependent on further technological advances and likely to incur appreciably higher costs.

39 See DECC (n 2) paras. 2.5–12; also, Climate Change Act 2008, pt I.

40 Cf no statutory commitment was included on the introduction of the Energy Bill (2012–13), save for a planned review of emissions objectives 2020–30 by 2016 (by reference to performance elsewhere in the EU).

41 DECC, *Planning our Electric Future: A White Paper for Secure, Affordable and Low-Carbon Electricity* (The Stationery Office 2011) 5–6.

42 DECC, *Overarching National Policy Statement (NPS) for Energy (EN-1)* (The Stationery Office, July 2011) pt 3.

43 Dir 2001/80/EC on the limitation of emission of certain pollutants into the air from large combustion plants [2001] OJ L 309/1 (27 November).

44 Dir 2010/75/EU on industrial emissions (integrated pollution prevention and control) (24 November).

decommissioning of (all but one of) the current generation of nuclear power stations.⁴⁵ The value of CCS to the task of capacity replacement can be envisaged from industry projections which suggest that (assuming proven technologically) around 20–30 GW could be CCS-linked by 2030.⁴⁶ Government seems similarly ambitious, seeking a total of 40–70 GW of new low-carbon generating capacity over that same period,⁴⁷ at least half of the median projection seemingly dependent on CCS deployment. Absent the unlikely achievement of reductions in net demand approaching an equivalent scale (for instance, through energy efficiency),⁴⁸ CCS emerges as a significant factor in securing longer-term compatibility of energy and climate change measures.

3 A governance perspective on regulating CCS

The UN's Intergovernmental Panel on Climate Change (IPCC) has been at the forefront both of the broad consensus over anthropogenic causes of global warming⁴⁹ and concerning the potential mitigatory role of CCS.⁵⁰ A growing cross-disciplinary engagement with climate change threats in the social sciences has also gained in political traction,⁵¹ with wide-ranging law and policy initiatives across numerous jurisdictions,⁵² although an ambivalent picture persists.⁵³ Reliance on CCS may, in this light, prove a double-edged sword, if resulting in stasis, or 'business as usual', so deflecting attention away from other laws and policies that can otherwise 'facilitate processes, products, and patterns of behaviour which are good for the planet'.⁵⁴ Within that wider picture, conflicting societal pressures have combined with a widespread inability among decision-makers consistently to prioritise low-carbon objectives,⁵⁵ vividly exemplifying a new form of disconnect between social thought and social action.⁵⁶ Legal responses to CCS can perhaps most advantageously be viewed from top-down perspectives, although it is interesting to consider first the relevance of a bottom-up approach to challenges posed.

The task of securing community buy-in, as part of the broader framework of climate change policy-making, raises participatory issues that are distinct from those seen more

45 See DECC, *NPS for Nuclear Power Generation (EN-6)* (The Stationery Office, July 2011) para 2.2; private sector resistance may likewise be leading to a new era of nuclear subsidy: D Carrington, 'Ministers Offer Nuclear Reactor Deal until 2050', *Guardian* (London, 19 February 2013).

46 CCS Association, *A Strategy for CCS in the UK and Beyond* (CCSA 2011).

47 DECC, *Carbon Plan* (2011) <www.decc.gov.uk/en/content/cms/tackling/carbon_plan/carbon_plan.aspx>.

48 Demand-side measures remain less of a government priority: see criticism by Energy and Climate Change Committee, *Draft Energy Bill: Pre-legislative Scrutiny* (First Report, HC 2012–13, 275–1, 23 July 2012).

49 IPCC, *Climate Change 2007: The Physical Science Basis* (Fourth Assessment Report, Working Group I, CUP 2007).

50 Metz et al (n 16).

51 For a critique of traditional economic perspectives on the processes for discounting future interests, see N Stern, *Economic Impacts of Climate Change: The Stern Review* (Cabinet Office/HM Treasury 2006).

52 See J de Cendra de Larragan, 'EU Climate and Energy Law: Challenges for Member States', in Peeters et al (n 7) ch 2.

53 T Jackson, *Prosperity without Growth: The Transition to a Sustainable Economy* (Sustainable Development Commission 2009).

54 D R Boyd, 'Sustainability Law: (R)Evolutionary Directions for the Future of Environmental Law' [2004] *Journal of Environmental Law and Practice* 357, 365.

55 In light of the major uncertainties and costs that surround CCS, it has been argued that as a viable part of the 'carbon mitigation toolkit' it is substantially undermined by multisectoral deficits in what has been termed the 'issue-attention cycle': see S Shackley and B Evar, 'Up and Down with CCS: The Issue-Attention Cycle and the Political Dynamics of Decarbonisation' in N Markusson, S Shackley and B Evar (eds), *The Social Dynamics of Carbon Capture and Storage: Understanding CCS Representations, Governance and Innovation* (Earthscan/Routledge 2012).

56 C D Stone, 'Response to Commentators (Should Trees Have Standing?)' (2012) 3 *Journal of Human Rights and the Environment* 100, 114; also illustrating the condition of bounded rationality: see eg A Green, 'Self Control, Individual Choice, and Climate Change' (2008) 26 *Virginia Environmental Law Journal* 77.

generally in environmental protection. For the UK, the emphasis on access to information and procedural issues,⁵⁷ shown to most significant effect in land use planning,⁵⁸ may at first sight have only limited connection with CCS. For instance, local attributes that typically characterise community resistance are largely missing from a scheme based upon undersea storage.⁵⁹ Yet two interrelated features point to the need to engage wider civil society in CCS: opportunities for participation can improve levels of engagement with environmental strategies more generally,⁶⁰ whilst at the same time they can enhance the processes of legitimisation that can otherwise seem out of place in the face of scientific or technological complexity,⁶¹ a condition surely encompassing CCS. Evidence as to how this can be achieved can be seen in steps taken towards developing a Scottish government communication strategy, specifically for CCS.⁶² This works from the premise that efforts to secure public acceptance, through mediation and participatory mechanisms, are a precondition to effective application.⁶³

Such efforts can be seen as supplementing wider discourse concerning reform of environmental governance, incorporating arguments for 'participatory dialogue and deliberation, devolved decision-making, flexibility . . . inclusiveness, transparency and institutionalised consensus-building'.⁶⁴ Similarly, it has been argued (in a discussion of 'social shaping' of CCS in the Australian context) that the challenge for regulators 'is not only to structure regulatory regimes to facilitate more comprehensive engagement processes, but also to find a way to support proponents with a range of benefits, should they do this well'.⁶⁵ There is also the wider European canvas on which participation commitments are mandated under the Aarhus Convention,⁶⁶ applicable EU Directives⁶⁷ and related national regimes (as in planning and strategic and project-based environmental assessment).⁶⁸ Experience of these more general measures can inform debates over specific risk-based strategies to address benefits and costs of deploying CCS as a primary vehicle for securing emissions reductions.⁶⁹

Turning to governance challenges more appropriately viewed from a top-down perspective, two illustrations are considered in the following discussion: first, the

57 W Howarth, 'Aspirations and Realities under the Water Framework Directive: Proceduralisation, Participation and Practicalities' (2009) 21 JEL 391.

58 S Owens and R Cowell, *Land and Limits: Interpreting Sustainability in the Planning Process* (Routledge 2002).

59 Cf B E Olsen, 'Wind Energy and Local Acceptance: How to Get Beyond the Nimby Effect' (2010) 19 *European Energy and Environmental Law Review* 239.

60 J Glicken, 'Getting Stakeholder Participation "Right": a Discussion of Participatory Processes and Possible Pitfalls' (2000) 3 *Environmental Science and Policy* 305.

61 J Steele, 'Participation and Deliberation in Environmental Decision-Making: Exploring a Problem-Solving Approach' (2001) 21 *Oxford Journal of Legal Studies* 415, 426.

62 J Hammond and S Shackley, *Towards a Public Communication Engagement Strategy for Carbon Dioxide Capture and Storage Projects in Scotland* (SCCS Working Paper 2010).

63 Ibid 13; cf the UK government's less particularized approach: DECC (n 2) 6.3.

64 Gunningham (n 21) 203.

65 P Ashworth and C Cormick, 'Enabling the Social Shaping of CCS Technology', in Havercroft et al (n 7) 259.

66 Aarhus Convention on Access to Information, Public Participation in Decision-making and Access to Justice in Environmental Matters (UN, 1998).

67 Especially Dir 2003/4/EC on public access to environmental information; Dir 2003/35/EC providing for public participation in respect of the drawing-up of certain plans and programmes relating to the environment.

68 For an analysis of the cross-applicability of existing UK measures to CCS, see M Lewis and N Westaway, 'Public Participation in UK CCS Planning and Consent Procedures', in Havercroft et al (n 7) ch 19; see also M Grecos, 'Carbon Capture and Storage: An Environmental Assessment' [2010] *JPL* 8, 12–15.

69 For a US perspective, see eg Shilling (n 32) 31.

implications of the limited degree of international law engagement in CCS; and, secondly, the burgeoning field of UK and EU climate change law and policy activity, highlighting the impact of CCS expectations upon wider concerns for energy sourcing. From the perspective of international law, Bugge has pointed out that legal issues of transboundary CO₂ chains under CCS are undertheorised.⁷⁰ Furthermore, the multilateral climate regime established under the Framework Convention on Climate Change⁷¹ and original arrangements under the Kyoto Protocol, which (to 2012) placed selective obligations on states for emissions reductions, provide no formal rules specifically referable to CCS.⁷² There has been no substantial impetus towards securing formal international agreement over CCS, in contrast with other areas of transnational risk. This perhaps reflects a shared, relatively sanguine approach to environmental risk in the CCS context, in notable contrast with the concern, say, for transboundary harms from nuclear accidents, referred to below. Nevertheless specific threats from, say, leakage from CCS storage sites add to an already fragmented picture of international law,⁷³ just as climate change impacts more generally threaten underlying objectives of a range of multilateral environmental agreements,⁷⁴ such as the Convention on Biological Diversity.⁷⁵

Indeed, attempts at producing geo-engineering solutions to climate threats raise novel issues of international law,⁷⁶ not least because there has hitherto been limited experience of a need 'to address large-scale human efforts to alter the environment'.⁷⁷ Short of any framework measure, CCS will fall to be taken into account on a case-by-case basis under discrete international regimes. The most significant example of this effect to date relates to controls placed upon dumping of waste materials. A prohibition of offshore sequestration under CCS, which had been contained in the 1996 London Protocol⁷⁸ to the 1972 Convention on the Prevention of Marine Pollution by Dumping of Wastes and Other Matter,⁷⁹ was lifted in 2007, subject to conditions based upon risk assessment and a management framework.⁸⁰ The effect is limited to permitting storage of captured CO₂ under the seabed and there remain therefore general concerns for 'international

70 For a discussion of uncertainties regarding allocation of state responsibilities across capture, transport and storage cycles, see H C Brugge, 'Transboundary Chains for Carbon Capture and Storage: Allocation under the Climate Regime between the States Parties of Emissions Due to Leakage' in Havercroft et al (n 7).

71 31 ILM (1992) 851.

72 Though it was confirmed at the conference of parties, Cancun (2010), that CCS may qualify in principle as a project activity under Kyoto's Clean Development Mechanism: Decision 7/CMP.6 <www.unfccc.int/documentation/decisions/items>; see further Brugge (n 70).

73 H Van Asselt, F Sindico and M A Mehling, 'Global Climate Change and the Fragmentation of International Law' (2008) 30 Law and Policy 423.

74 See C Redgwell, 'Climate Change and International Environmental Law' in R Rayfuse and S V Scott (eds), *International Law in the Era of Climate Change* (Edward Elgar Publishing 2012).

75 1670 UNTS 79.

76 See C Redgwell, 'Geoengineering the Climate: Technological Solutions to Mitigation Failure or Continuing Carbon Addiction?' (2011) 2 Carbon and Climate Law Review 178.

77 Redgwell (n 74) 138–40, referring exceptionally to the Convention on the Prohibition of Military or Any Other Hostile Use of Environmental Modification Techniques, which suggests in some respects a model which might be modified for civil application.

78 36 ELM 1 (in force March 2006).

79 11 ILM 1294.

80 See R G Lee, 'Sub-seabed Carbon Sequestration: Building the Legal Platform' (2009) 30 Liverpool Law Review 131; also, C Redgwell, 'International Legal Responses to the Challenges of a Lower Carbon Future: Energy Law for the Twenty-First Century' in S Farrell, T Ahmed and D French, *Criminological and Legal Consequences of Climate Change* (Hart Publishing 2012) 42–45.

responsibility and liability for leakage and other damage to the marine environment',⁸¹ which can be linked, as seen above, with the lack of movement toward international agreement.⁸²

The second illustration, which raises issues as to energy choices at national (and regional) levels, relates to the sourcing of energy and the danger of unsustainable 'lock-in'.⁸³ The lack of experience of large-scale CCS presents the energy sector with a temporal quandary. A consensus view suggests that the earliest that deployment might be expected to begin would be around 2030.⁸⁴ The installation of CCS infrastructure and systems to anything approaching sufficient scale might not then realistically be expected until around mid-century.⁸⁵ This poses a more fundamental problem in light of the close links between future expectations for CCS and levels of fossil fuel use in the meantime, the timescale standing in the way of a shift away from non-sustainable energy sources.⁸⁶ The effect might prove temporary, as an essential means of securing the transition to a low-carbon future, with CCS seen as an essential 'bridging technology' towards climate change mitigation.⁸⁷ On the other hand, if the effect were to reinforce the impact of system inertia under settled institutional arrangements (public and private),⁸⁸ the biosphere may be brought closer to what climate scientists term a tipping-point.⁸⁹

Major consequences follow: first, in the coming decades, and before significant levels of CCS can be put in place, the extension globally in fossil fuel-generating capacity will be less likely to abate,⁹⁰ with linked infrastructure investment expected to have a lifespan in excess of 40 years. To counter this, some regional regimes, applying the more direct coercive approaches of command-and-control, may insist upon 'carbon capture ready' installations. Under EU legislation, whilst CCS deployment is not mandatory,⁹¹ member states are now required to assess new projects for this purpose,⁹² and (where capture is feasible) to provide suitable site capacity for that future event. Thus, subject to overall emission limits set under

81 R Rayfuse, 'Climate Change and the Law of the Sea' in Rayfuse and Scott (n 74) 166–69.

82 Informally, a Clean Energy Ministerial of participating states shares lessons and best practice through a Carbon Capture Use and Storage Action Group: <<http://cleranenergyministerial.org/events/cem3/index.html>>.

83 P J Vergragt, N Markusson and H Karlsson, 'Carbon Capture and Storage, Bioenergy with Carbon Capture and Storage, and the Escape from the Fossil-Fuel Lock-in' (2011) 21 *Global Environmental Change* 282.

84 See, for instance, UN Development Programme, *Avoiding Dangerous Climate Change: Strategies for Mitigation* (Human Development Report, UN 2007/08) 145–46.

85 E Rubin, 'Technical Summary', in Metz et al (n 16) 41. The IEA suggests that respectively 100 and 3000 large-scale plants are required to be in operation by 2020 and 2050.

86 G C Unruh, 'Understanding Carbon Lock-in' (2000) 28 *Energy Policy* 817.

87 Dir 2009/31/EC (n 24) preamble, para 4; see S Shackley and M Thompson, 'Lost in the Mix: Will the Technologies of Carbon Dioxide Carbon Capture and Storage Provide Us with a Breathing Space as We Strive to Make the Transition from Fossil Fuels to Renewables?' (2012) 110 *Climatic Change* 101.

88 Both privatized, free market and public, regulatory approaches are capable of imposing externalities, in effect applying discount rates upon the protection of future environmental interests: see D H Cole, 'Clearing the Air: Four Propositions about Property Rights and Environmental Protection' (1999) 10 *Duke Environmental Law and Policy Forum* 103, 117–25.

89 See J Hansen, 'A Slippery Slope: How Much Global Warming Constitutes "Dangerous Anthropogenic Interference"?' (2005) 68 *Climatic Change* 269.

90 See eg Johansson et al (n 8).

91 Dir 2009/31/EC (n 24) art 4; this reflects such factors as continuing uncertainty as to when the technology will be available for successful deployment, disparities in geological features, as well as in evidence of public concern.

92 Applicable to proposals for new 300MW-plus combustion plants, requiring assessment as to the availability of storage sites and feasibility of transport facilities: *ibid* art 33.

the Emissions Trading Directive (EU emissions trading scheme (ETS)),⁹³ such installations can continue to be built and permitted for electricity generation, lack of CCS notwithstanding. This might be ameliorated if EU-wide energy performance standards were imposed, effectively ruling out certain forms of generation until CCS becomes available.⁹⁴ In the UK's case, its 2012–13 Energy Bill contains proposals for an energy performance standard, which, absent CCS, would effectively rule out further substantial investment in coal-fired power stations.⁹⁵

Secondly, the expectation of CCS offers incentives to decision-makers to scale down commitments to low-carbon investment.⁹⁶ The significance of this is comparable with that arising from so-called carbon leakage, produced by outsourcing of industrial production (and thus carbon emissions) to developing economies.⁹⁷ This has led to calls for 'border tariffs' on carbon intensive imports, so as to shift focus onto 'consumption of embodied carbon dioxide emissions, rather than production'.⁹⁸ But an important consequence of the above is that private investment decisions, faced with the resulting uncertainties, will respond accordingly, a logic which still applies with considerable force in the energy markets.⁹⁹ So long as market-makers are confident of sourcing fossil fuels at competitive prices, then in default of further regulatory constraints this can justify further diversion of resources away from alternatives.

4 CCS and re-regulating the energy system

The task of addressing those financial risks that determine investment decisions is thus a key factor in framing regulatory terms for CCS. Investment obstacles arise across a range of timeframes, but an early test relates to securing seed-corn funding in advance of capacity to make informed judgments affecting CCS for the future. Significant cost exposures accrue through project lifecycles. These arguably place CCS beyond the commercial capabilities of both the industries directly concerned (in view of the costs of financing such ventures and the shorter-term returns expected by shareholders) and those capital markets that support investment (for instance, through equity or loan stock issues, or other forms of leveraged finance). For early-movers, the vexed question of start-up costs exemplifies a form of 'chicken and egg syndrome', on account of the lack from the outset of the kind of reliable performance data that is normally a prerequisite to investment decisions.¹⁰⁰ In these

93 Dir 2003/87/EC (n 24); Decision No 406/2009/EC on the effort of member states to reduce their greenhouse gas emissions to meet the Community's greenhouse gas emission reduction commitments up to 2020 [2009] OJ L140/136.

94 Following review of CCS readiness that the European Commission is required to carry out (in 2015): see Dir 2009/31/EC (n 24) art 38.

95 Under cl 38, maximum annual emissions from all new plant (over 50MW) would be limited to 450g CO₂/KWh; rather than (ironically) 'undermining decarbonisation', CCS plants were subject to the same restriction after the draft Bill stage: *Government Response to HC Energy and Climate Change Committee Report on the Draft Energy Bill* (Cm 8504 November 2012) 30–21.

96 Greenpeace International, *False Hope: Why Carbon Capture and Storage Won't Save the Climate* (Greenpeace 2008) pt 4.

97 W Straw, D Nash and R Balfour, *Europe's Next Economy: The Benefits of and Barriers to Low-Carbon Transition* (Institute for Public Policy Research, May 2012).

98 See C Hepburn, 'The Energy Mix, Carbon Pricing and Border Carbon Adjustments' (2012) 24 *Environmental Law and Management* 177, 181; also, D Helm, C Hepburn and G Ruta, 'Trade, Climate Change and The Political Game Theory of Border Carbon Adjustments' (2012) 28 *Oxford Review of Economic Policy* 368.

99 See D Helm, 'Climate Change Policy: Why Has So Little Been Achieved?' (2008) 24 *Oxford Review of Economic Policy* 211.

100 J P Tomain, *Ending Dirty Energy Policy: Prelude to Climate Change* (CUP 2012) 113.

circumstances the logic of the market would suggest that investment can be more advantageously directed elsewhere.

In view of the potential benefits from CCS, threats to commercial viability have significant implications for the public sphere,¹⁰¹ and hence the broad acknowledgment of the need for public investment and state support towards costs especially of early demonstration projects. Yet, in the UK the government's 'commercialisation programme' has met considerable private sector resistance, reflecting (as well as restrictive terms on offer) an industry belief that proposals fall short of what is needed.¹⁰² Likewise, though for different reasons, EU supplementary support for CCS has also been called into question. Its funding, reliant on allocations from the EU ETS 'new entrants reserve',¹⁰³ has been beset by delay, and is handicapped by depressed ETS carbon allowance prices, eroding subsidy levels.¹⁰⁴ Nevertheless the link between CCS and the ETS for funding purposes creates a valuable precedent. Early phases of ETS supported grandfathering through free distribution of allowances. From 2013 auctioning is being introduced for some sectors, so the capability arguably exists (despite fears of carbon leakage as costs increase) for generating a funding base which might be specifically earmarked for future CCS and other low-carbon initiatives.

Looking toward operational phases, both economic and environmental costs can be expected to be significant. These include threats to sustainability from a growing need to 'allocate more and more energy just to obtain energy', and thus 'declines in energy returned in energy invested'.¹⁰⁵ For instance, it has been estimated that retro-fitting CCS capacity to existing plants could result in a 50 per cent increase in capital costs.¹⁰⁶ So-called 'energy penalties' are also of potential significance, representing efficiency losses through the extra energy consumption that CCS requires.¹⁰⁷ Further overall costs to the generator are variously assessed, from around a 20 per cent increase to a near doubling,¹⁰⁸ at the higher end, especially on expensive coal-firing retro-fit.¹⁰⁹ Moreover, though much the most significant, such costs largely relate to capture alone, whereas investors must take account of further costs at remaining CCS stages (infrastructure itself, transport to storage sites, maintenance and monitoring thereafter). Thus (especially with large-scale CCS not yet

101 W B Jacobs, L Cohen, L Kostakidis-Lianos and S Rundell, *Proposed Roadmap for Overcoming Legal and Financial Obstacles to CCS* (Discussion Paper 2009-04, John F Kennedy School of Government, Harvard 2009) 18–22, proposing an 'equitable approach' based on public-private cost-sharing: with federal support by capping liabilities for 'earliest movers' (demonstration projects); and subsequently for all sequestration site operators, in excess of any funds contributed by the industry *ex ante* (financial assurance or risk pooling).

102 With £1 billion demonstration project subsidy through competitive process; subsequently relaunched on new terms: see *ENDS Report* 447 (London, April 2012) 34–35; *ENDS Report* 448 (London, May 2012) 12.

103 The 'NER 300' project is based on 300 million allowances available (until 31 December 2015), for up to 10–12 demonstration projects (with a maximum for any one of 45 million allowances).

104 'CCS at Risk Following Carbon Price Collapse', *ENDS Report* (London, 16 November 2011); also, 'CCS and Renewables Plans Hit by EU Finance Blow', *ENDS Report* (London, 12 July 2012); 'Cash for Carbon Capture Projects Slashed by Eurozone Crisis', *Guardian* (London, 6 August 2012).

105 See J P Manno, 'Less Energy, Better Health' in L Westra, L Soskolme and D W Spady, *Human Health and Ecological Integrity: Ethics, Law and Human Rights* (Earthscan/Routledge 2012) 83–86 (analogy with the action of entropy under Second Law of Thermodynamics seems clear).

106 McKinsey & Co, *Carbon Capture and Storage: Assessing the Economics* (Report, McKinsey & Co 2008) 10

107 Estimates are for between 10% and 40% of electricity produced being lost in this way: see Metz et al (n 16) 3.

108 Rubin (n 85) 40–43.

109 Massachusetts Institute of Technology, *The Future of Coal* (MIT 2007) 29 <http://web.mit.edu/coal/The_Future_of_Coal.pdf>.

proven) doubts over prospects for cost-competitiveness remain key determinants in energy market responses.¹¹⁰

It is therefore relevant to explore aspects of the wider relationship between the state and energy markets. There has long subsisted a curious disconnect between energy and the environment in policy-making and legal frameworks.¹¹¹ Indeed, traditionally, discrete approaches and priorities have tended to defy the capacity for environmental law in other contexts to integrate into other fields.¹¹² As recently as 1995, echoing a prevailing deregulatory agenda since the 1980s that has affected the energy markets as elsewhere,¹¹³ the UK government was persisting in seeing ‘no reasons why the electricity market should not of its own accord provide an appropriate level of diversity’.¹¹⁴ Yet obstacles in the way of maximising low-carbon energy investment require an energy policy framework, especially where there is a perceived need for incentives.¹¹⁵ Dual objectives for climate change mitigation and energy security have become a catalyst for a departure from the hands-off approach, with current legislative proposals being described as pointing towards a form of ‘re-regulation of the energy system’.¹¹⁶ Similarly, albeit that the EU enjoys more limited competences in respect to energy as compared with the environment, the European Commission has acknowledged the need for more effective policy responses, for ‘energy and the environment can no longer be viewed in isolation’.¹¹⁷ Thus efforts towards meeting objectives of security, diversity and sustainability are creating a renewed focus on directed energy planning, albeit superimposed onto the liberalised market structure,¹¹⁸ of which the Energy Bill is a complex manifestation.

New forms of selective state interference in energy markets can be illustrated, first, in legislative proposals for ‘electricity market reform’.¹¹⁹ These include an overhauling of public incentives for low-carbon investment, replacing a mechanism based on a somewhat hit-or-miss feed-in tariffs (FiT) procedure,¹²⁰ by offering fixed, differential payment guarantees (for up to 25 years),¹²¹ also available for utilisation of CCS. The idea in principle

110 DECC (n 2) 4.2.

111 Not a UK phenomenon alone: see, for instance, L D Davies, ‘Alternative Energy and the Energy–Environment Disconnect’ (2010) 46 *Idaho Law Review* 473.

112 Cf N Dhondt, *Integration of Environmental Protection into Other EC Policies* (Europa Law Publishing 2003).

113 One incident arguably being a former moratorium on new nuclear investment (expensive both in set-up costs, long-term risks and potential liabilities), more recently superseded by proposals for a new generation of nuclear power stations; in the meantime regulators have approved extended terms in the useful lives of some installations still in operation: see DECC (n 45).

114 Department of Trade and Industry (DTI), *The Prospects for Nuclear Power in the UK: Conclusions of the Government’s Nuclear Review* (Cm 2860, 1995) 2.7–8; at that time concluding that no near-term justification existed for a new generation of nuclear power stations.

115 See, for instance, M Pallemerts, ‘Climate Change, Natural Gas and the Rebirth of Energy Policy’ in P Winand (ed), *Securing Sustainable Energy Supplies in Europe and Australia* (Pieter Lang 2011).

116 Energy and Climate Change Select Committee, *Draft Energy Bill: Pre-legislative Scrutiny* (First Report, HC 2012–13 275–1 July 2012) ch 9.

117 European Commission, *An Energy Policy for the European Union* (White Paper COM(95) 682 final, 13 December 1995) para 128.

118 A significant milestone for the new era of energy planning was the DTI, *Conclusions of the Review of Energy Sources from Power Generation* (HMSO 1998).

119 Energy Bill 2012–13, pt 1.

120 Under Energy Act 2008, ss 41–43, which arguably led to an undue advantage for solar voltaic installations; an attempt by the minister to reduce subsidy levels with partly retrospective effect was struck down in *R (Homesun Holdings Ltd) v Secretary of State for Energy and Climate Change* [2011] EWHC 3575.

121 Under a Renewables Obligation, electricity suppliers have been required to purchase (increasing proportions of) generated power from renewable sources, and low-carbon generators receive market prices topped up by guaranteed payments under tradable Renewables Obligation Certificates.

secures two advantages: greater certainty for investors,¹²² under long-term ‘contracts for difference’ (FiT CfD),¹²³ alongside greater protection for the taxpayer.¹²⁴ Low-carbon investment would thereby be cushioned from carbon price collapses (so protecting market competitiveness against fossil fuel generation) whilst potential windfalls from higher carbon prices could be subject to claw-back. Secondly, as previously seen, an emissions performance standard is proposed,¹²⁵ effectively ruling out non-CCS abated coal-fired stations;¹²⁶ co-existing with a separate requirement that all new fossil fuel (in effect, gas-fired) power stations be ‘carbon capture ready’.¹²⁷ A third, related mechanism, focuses on pricing, seeking a more consistent incentivising of low-carbon technology through the introduction of a national carbon price floor (from April 2013).¹²⁸ This would limit the impact of depressed price levels for the allowances that must be purchased by installations that exceed allotted caps under the EU ETS scheme.¹²⁹

Interventions such as through FiT CfD depend on a level of support that produces ‘greater revenue certainty . . . [which] should lead to developers being able to reduce the costs of financing their investments’.¹³⁰ The specific inclusion of CCS is aimed at facilitating its development ‘into a mature technology capable of being assessed by investors on a normal commercial basis’.¹³¹ Although the consequence would be to secure reduced carbon intensity of the energy system alongside increased energy security, the operation of the markets will continue to exert a major influence on outcomes. An emerging feature, for instance, with coal in the UK and much of the rest of the EU having been in decline (perhaps to be invigorated depending on events affecting CCS), has seen prospects for gas on a strongly upward curve. Despite extensive price increases in recent times, this arguably reflects continuing industry concerns at levels of low-carbon investment, as well as increased confidence in levels of global gas supply.¹³² Here fundamental tensions are apparent as to the implications of a new ‘dash-for-gas’, increasing:

122 DECC (n 47) 3.53, following the Climate Change Committee’s recommendations for the fourth carbon budget (up to 2027), in accord with the obligations under Climate Change Act 2008, ss 12, 14.

123 Energy Bill 2012–13, cls 2–16.

124 It is the intention that FiT CfD replaces the Renewables Obligation, the former financial support mechanism incentivising renewable energy projects, by 2017.

125 Absent EU action likewise, European Commission approval would be required to allay the risk of legal challenge: see L Squintani, M Holwerda and K de Graaf, ‘Regulating Greenhouse Gas Emissions from EU ETS Installations: What Room is Left for Member States?’ in Peeters et al (n 7).

126 Energy Bill 2012–13, cl 38(1)(2); though not ruling out gas (also a fossil fuel, but around 60% more carbon efficient).

127 The Bill enables ministerial regulations with the effect of disregarding emissions from such plants using CCS technology for the purpose of meeting emissions limits: cl 38(6)(a), (7)(a)(b)(c).

128 Although market circumstances differ, the minimum guideline price appears not to be significantly out of line with the conclusions of a 2007 report by MIT, which suggested that the carbon price should not fall below \$25 per tonne: see J Deutch and E J Moniz, *The Future of Coal: An Interdisciplinary Study* <<http://web.mit.edu/coal/>>.

129 Attempts to impose EU-wide measures to support ETS market prices have so far been unsuccessful: see D Carrington, ‘Blow to Green Hopes after Price of Carbon Permits Fall to Record Low’, *Guardian* (London, 25 January 2013). The UK’s national scheme would be introduced by adjustments in its domestic climate change levy.

130 DECC (n 2) 4.6.

131 *Ibid* 4.3.

132 IEA, *World Energy Outlook 2012* <www.iea.org/publications/freepublications/publication/English.pdf>.

the risk of locking the UK into a high-carbon electricity system and represents a huge gamble on the eventual availability of cost effective [CCS] technology for gas plants. This could pose a severe threat to the achievement of our long term climate change goals.¹³³

There are therefore considerable obstacles in the way of intervention efforts to nudge the energy economy onto a different trajectory by assisting low-carbon generators better to absorb price risks,¹³⁴ without disabling the confidence of those whose energy choices remain largely constrained by the market.¹³⁵ Outside of the advantages of a free market, its *raison d'être* does not lie in the delivery of regulatory goals. Indeed, there is a propensity if unchecked to externalise costs, especially in the face of unknowable risk; a characteristic of problems posed by climate change generally, and in some respects also by CCS. Such consequences for low-carbon energy can be witnessed as above in market signals that favour renewed reliance on gas, and there remains a risk that these electricity market reforms will perpetuate a 'stop-start approach' to low-carbon investment.¹³⁶

With effective regulatory frameworks crucial determinants towards successful deployment of CCS,¹³⁷ experience of early development projects will afford opportunities to test aspects of suitability.¹³⁸ Regulatory responses must address inherent risks from failures across the whole capture-transportation-storage cycle, alongside new dilemmas from uncertainties attaching to potential liabilities. Before exploring further, some comment is necessary regarding the wider EU legal context, where, as in the UK, we are witnessing a growing interdependence across fields of energy and climate change. New legislative incursions are in the course of development, following the catalysing effect of the 2007 Lisbon Treaty, which introduced an autonomous energy title,¹³⁹ contrasting with a previous legal model which failed to address energy on its own account.¹⁴⁰ Whilst EU constraints are increasingly influential,¹⁴¹ widespread national autonomy persists as to the detail of climate change policy; and similarly energy policy is still largely determined at member state level,¹⁴² with, for

133 Energy and Climate Change Select Committee, *UK Energy Supply: Security or Independence?* (Eighth Report 2010–12, HC 1065, October 2011), para 103.

134 In August 2012 DECC issued a call for evidence on barriers to securing long-term power purchase agreements.

135 Delays in provision of EU and UK financial support (already affecting the competitive process for demonstration projects and possibly also uncertainties as to longer-term guarantees or assumptions of liability) have produced so called 'investor fatigue' as proposals have been withdrawn and major hydrocarbon companies vacated the market: see DECC Science Advisory Group, *Briefing Note on Carbon Capture and Storage* (DECC, July 2011) 10–11.

136 See *ENDS Report* 452 (London, September 2012) 5–6.

137 Sites in principle in the UK are all offshore and include depleted hydrocarbon reservoirs, active fields with opportunities for enhanced hydrocarbon recovery, and deep saline aquifers: see DECC (n 2) 40, for a study of UK storage capacity, which is estimated to be enough for 100 years of current use, representing less than a quarter of potential storage capacity.

138 Following a mock CCS project the Scottish government and Scottish Centre for Carbon Storage have already produced a regulatory test toolkit, commended by the European Commission as a model of best practice: <www.scotland.gov.uk/Topics/Business-Industry/Energy/resources/Publications/CCSRegulatoryToolkit>.

139 Art 194 TFEU; expressed to be without prejudice to law-making powers under the Treaty's environmental title (art 191).

140 Early fruit of which shift can be seen in Dir 2012/27/EU on energy efficiency [2012] OJ L315/1 (14 November).

141 Eg EU ETS Dir 2003/87/EC (n 24); also Effort-Sharing Decision No 406/2009/EC (n 93).

142 See further Pallemarts (n 115).

instance, the energy title explicitly reserving national sourcing of energy supply.¹⁴³ The EU is further limited in its choice of instruments, a prime illustration of which has been its inability to make any headway in extending its competence in relation to green fiscal measures.¹⁴⁴

5 Future regulatory challenges for CCS

The choice and terms of suitable CCS regulatory mechanisms appear analogous to those used by the EU elsewhere in response to climate change.¹⁴⁵ This broadly relies on utilising command-and-control methods alongside market-based alternatives, as exemplified by efforts through the EU ETS to encourage CCS by express inclusion in its scope. Here we see a paradigm illustration of markets in non-directed economies as both the prime vehicle for resource exploitation, as well as the basis for a shift toward economic mechanisms which nevertheless create further complex problems of regulation.¹⁴⁶ In principle, captured carbon volumes under CCS are not classified as emissions, but compliance with mandated mitigation commitments being at stake, any subsequent leakages would be transformed into non-permitted emissions by a process of accounting adjustment.¹⁴⁷ A cost would then be imposed by the triggering of a requirement to surrender emissions allowances equivalent to the estimated extent of leakage. Such a deterrence strategy, however, is vulnerable to low carbon prices,¹⁴⁸ and a continuing failure under ETS to assure realistic pricing (to shift the balance of investment advantage away from non-mitigated fossil fuels emissions) constitutes a significant constraint on development of CCS.¹⁴⁹

For the most part, however, the EU's CCS framework can be described as premised upon command-and-control, primarily founded on a permitting regime.¹⁵⁰ The new regulatory focus is upon the feature of storage, including site selection, risk management and closure. A pervasive uncertainty relates to risks of harm from migration from storage sites: including, as seen above, to climate change mitigation goals, but also harm to human health or property and to the wider environment. Risks can arise in ways that are novel in their application to CCS, especially related to long-term storage, with potentially open-ended liabilities. They can also arise in ways that are incidental to existing public law regimes, or under established private law causes of action. In the latter respect, relevant harms are likely to be more localised, and for UK purposes would include common law rules for establishing liability for harm to persons and property. From a health perspective, for instance, CO₂, whilst unlike oxygen not explosive or otherwise flammable, is corrosive, and leakage in confined spaces carries a threat of asphyxiation even at concentration levels of

143 The 2012 disaster at Fukushima has led to contrasting national responses: the UK remains formally committed to a new generation of nuclear power stations, whereas Germany, for instance, is now committed to phase them out.

144 A Jordan, et al, 'Governing with Multiple Policy Instruments?' in A Jordan and C Adelle (eds), *Environmental Policy in the EU* (3rd edn, Earthscan/Routledge 2013).

145 For a useful summary of measures differentiated by regulatory category, see *ibid* 317.

146 R Macrory, 'Regulating in a Risky Environment' [2001] *Current Legal Problems* 619.

147 See Dir 2009/31/EC (n 24) art 16.

148 A discouraging general trend persists: see n 104, charting a recent fall to €2.38 per tonne, far below levels required for stimulating investment (circa €20-plus). Moreover political obstacles have thus far stymied Commission attempts to bolster prices through a temporary withholding of allowances from the market: see F Harvey and A Vaughan, 'MEPS Reject Key Reform of Emissions Trading Scheme', *Guardian* (London, 16 April 2013).

149 Committee on Climate Change, *Meeting Carbon Budgets* (Third Progress Report to Parliament, Stationery Office, June 2011) 17–18.

150 Dir 2009/31/EC, arts 6–16 (including the permit regime, alongside monitoring, reporting and remedial measures); an exploration permit is required under art 5.

as low as 3 per cent by volume.¹⁵¹ However, for the UK (with only offshore storage currently envisaged), health risks are more likely to result from operational activities, including transportation, and are less likely to follow from migration from storage.

In relation to public law impacts, certain discrete environmental protection regimes have a potential for overlap with CCS regulation. In numerous instances, compliance with established regulatory arrangements protecting against pollution or harm to the wider environment will be required,¹⁵² including those relating to air pollution, soil contamination, surface or ground water pollution,¹⁵³ and damage to habitats. A significant area of cross-cutting regulation affects risks of harm to the wider environment and the mandatory requirements for corrective measures under the Environmental Liability Directive. This directive recognised extended notions of harm, each of which might be relevant to CCS: including land or soil contamination with serious risk to public health; water pollution involving reduced ecological or chemical quality status; or other environmental harms in the event of significant damage caused by defined dangerous activities that pose risk to humans and the environment (as regulated elsewhere under EU legislation and thus within EU legal competence).¹⁵⁴

By contrast, with CCS storage being of its nature intended to be permanent and not temporary,¹⁵⁵ the application of certain other regimes can be inappropriate. These include those conditional protections and exemptions for purposes of marine protection under specific arrangements for licensing of deposit of articles or materials on the sea bed in relation to construction activities.¹⁵⁶ Other domestic and EU regimes could in principle be in conflict with activities sanctioned by CCS permitting. These mainly relate to waste control and measures against dumping at sea (also, obligations under international law referred to previously), and it has been necessary to negotiate the terms on which exemptions might apply. Thus capture, transport and storage are expressly excluded from the definition of waste under the Waste Framework Directive.¹⁵⁷ By contrast, capture facilities are subject to public regulation in the same way as the parent-generating installations of which they form part,¹⁵⁸ and likewise transport of compressed CO₂ by pipeline is subject to legislation already in place for the oil and gas industry.¹⁵⁹

Apart from such differential regulatory applications, a distinctive feature of CCS relates to post-closure risks arising during the storage phase (that is, post-capture, transport and

151 It is denser than air and will pool in low-lying areas, posing hazards to those exposed where ventilation is poor: see further D Spreng et al, 'CO₂ Capture and Storage: Another Faustian Bargain?' (2007) 35 *Energy Policy* 850.

152 Eg art 31, amending Dir 85/337/EEC concerning environmental impact assessment of projects (such as capture, transport and storage).

153 Ocean acidification is a major reason why underwater dispersal is not seriously considered for CCS purposes, though carbon leakage from sites may yet pose future threats.

154 Strict liability applies. The directive further recognises specifically environmental claims for biodiversity damage (in the very specific sense of serious prejudice to the conservation status of EU habitats and wild birds directives), on the basis of a fault-based liability on account of any activities (whether or not otherwise regulated). The directive is constrained in numerous further respects: neither operating retroactively, nor extending to harm to personal interests in property or economic loss, any remedy for which remains a matter for national civil liability arrangements. It is moreover open to member states to apply state-of-the-art and regulatory compliance defences.

155 See also Water Resources Act 1991, s 88.

156 Food and Environmental Protection Act 1985, pt II.

157 Dir 2009/31/EC (n 24) art 35, amending Dir 2006/12/EC, art 2(1)(a); also art 36, excluding the shipment of waste under reg 1013/2006, art 1(3).

158 Electricity Act 1989, s 36.

159 Such as authority for onshore pipeline construction under the Pipelines Act 1962; see also Pipeline Works (Environmental Impact Assessment) Regulations 2000/1928.

injection).¹⁶⁰ With an indefinite potential for causing subsequent environmental and other harms, further questions of risk acceptability and commercial viability arise. In most jurisdictions addressing this, the notion of transfer of contingent liabilities to the state is emerging as a common feature of CCS regulation.¹⁶¹ This reflects a fundamental misalignment of risk: in light both of the mutability of corporate operators and the limits on levels of financial assurance (including insurance) that can realistically be negotiated in the market. So a key question relates to the terms on which risk transfer takes place.¹⁶² For instance, in Australia, liability passes by operation of law to the Commonwealth government (for storage offshore) once at least five years have passed post-closure, provided that the minister is satisfied that no significant risks are posed, and subject to a 15-year closure assurance period. Liability extends to 'any' liability, seemingly including even liabilities under the common law (and it is expressly provided that the Commonwealth is also liable where the holder ceases to exist). This contrasts with approaches adopted in the states of Victoria and Queensland (responsible for CCS licensing inland), under which licensed operators may retain liability exposure for an indefinite period.¹⁶³ The Commonwealth measure is arguably the more realistic, especially the longer that time elapses and given the above mutability of commercial entities.

Similarly, EU arrangements set out criteria which include the passage of a minimum post-closure period (which the authority may waive) of 20 years, the operator meeting financial obligations (requiring contribution to a financial mechanism to pay the authority's post-transfer costs), and the authority being satisfied that all available evidence indicates complete and permanent containment.¹⁶⁴ The approach, on the face of it a pragmatic compromise, seems to ride two horses in different directions: once approved, imposing effective long-term responsibility on the state,¹⁶⁵ though also allowing authorities with higher risk aversion to seek to avoid transfer risks, by adopting perhaps a zero-risk threshold, a notion elsewhere in EU jurisprudence expressly rejected for application of the precautionary principle.¹⁶⁶ However, a blanket refusal is likely to be met with administrative law challenge, and perhaps of greater practical significance an increased likelihood of corporate avoidance techniques (such as reduced self-reporting of problems). Collateral impacts on investor confidence in CCS within that jurisdiction might also follow, and so, as projects move towards the closure phase, the expectation must be that indefinite state liability should be anticipated as the likely outcome.

Exclusion of liability under the EU's transfer arrangements may nevertheless be disapplied in the event of operator fault. This notion is expressed in extremely broad terms: including wilful deceit, concealment of relevant information, failure to exercise due diligence,

160 Dir 2009/31/EC (n 24) contains general provisions including those for selection of storage sites, including exploration permits: arts 4, 5; storage permits, arts 6–11; and operations including closure and post-closure, including monitoring: arts 12–20.

161 IEA (n 1) 91–105.

162 Including as to the need for delay in transfer eg reflecting outstanding issues relating to earlier incidents.

163 M Gibbs, 'The Regulation of Geological Storage of Greenhouse Gases in Australia' in Havercroft et al (n 7) 167–70; also A Warburton, J Grove, S Then and K Geddes, 'Geosequestration Law in Australia' in T Bodyhandy and P Christoff (eds), *Climate Law in Australia* (Federation Press 2007); though in each state statutory ownership of stored gases appears to become state property, suggesting that the state is likely to be first defendant in any third-party action.

164 Dir 2009/31/EC (n 24) art 18(1).

165 Ibid, applicable to obligations under each of the CCS, Environmental Liability and ETS directives; although liability transfer would not apply to other regimes under EU law (such as integrated pollution prevention and control) or domestic law (such as statutory nuisance).

166 Case T-13/99 *Pfizer Animal Health SA v Council of EU* [2002] ECR II-3305.

negligence, or what is termed deficient data, in which event there appears to be a duty to recover post-transfer costs thereby incurred.¹⁶⁷ It may be envisaged that this unusual formulation regarding fault will create uncertainty, especially in the early period following transfer, where the notion of being 'at risk' will be a more realistic possibility. But in the post-operational phases of most economic activities, given a strong scientific research base, circumstances in which liability can arise can be expected thenceforth to reduce. Moreover, although the directive's notion of 'all legal obligations' is further constrained by competence limitations (so excluding private law obligations),¹⁶⁸ where a potential for harm persists, inherent limits apply as to how far a civil liability regime can capture such risks within rules applicable to recognised causes of action (including as to causation, remoteness and limitation of actions); and, in consequence, such losses would be likely to be left where they lie.

In concluding this section, the distinctiveness of CCS risk transfer is further considered by reference to two partly cognate regimes. First, contrasting regulatory responses may be drawn with regard to liability for nuclear accidents, where in an effort to counter uncertainty, both for entities active in the nuclear industry and respecting state concerns for the public good, two main conventions are placed.¹⁶⁹ Despite the lack of a comprehensive regime,¹⁷⁰ a shared feature is that exposure to liability is, in the first instance, channelled through the operator, with responsibilities thereafter assumed (in effect, guaranteed) by respective states.¹⁷¹ Operator liability provisions are circumscribed by reference to harm,¹⁷² financial limits¹⁷³ and limitation periods,¹⁷⁴ with flexibility under which insurance or other financial security arrangements for meeting maximum liabilities are reserved to state level.¹⁷⁵ UK operator liabilities are limited to a relatively low maximum exposure relative to other states.¹⁷⁶ Exposure to liability would significantly increase were the 2004 Protocol in force,¹⁷⁷ affecting both operators and states as

167 Dir 2009/31/EC (n 24) art 18.7.

168 Member states may exclude operators from such liabilities.

169 The first, under the Organization for Economic Cooperation and Development, through the 1960 Paris Convention on Third Party Liability in the Field of Nuclear Energy, 8 EYB 203 (as supplemented by the 1963 Brussels Convention, 2 ILM 685 (with a Protocol in 2004, not yet in force), including most Western European states; the second, under the International Atomic Energy Agency, through the 1963 Vienna Convention on Civil Liability for Nuclear Damage, 2 ILM 727, amended by protocol in 1997, 36 ILM 1462, mostly covering the former Soviet bloc.

170 A blueprint for one is contained in a 2007 Convention on Supplementary Compensation.

171 See eg Paris Convention (n 169) art 6.

172 Ibid art 3(a) includes harm to humans and damage to property.

173 Ibid art 7 (amended Brussels Convention, art 3) establishes tiers of liability and state discretion in setting for operators a minimum requirement (circa €5 million), up to a much increased set maximum (€300 million), reliance for up to half of which might come from a solidarity provision under art 12.

174 Ibid art 8.

175 Ibid art 10.

176 Nuclear Installations Act 1965, ss 7ff. The UK maximum is £140 million (Nuclear Installations (Increase in Operators' Limits of Liability) Order 1994/909. For comparative illustrations from other jurisdictions, see S Tromans, *Nuclear Law: The Law Applying to Nuclear Installations and Radioactive Substances in its Historic Context* (2nd edn, Hart Publishing 2010) 185–88.

177 Resistance to the Protocol reflects eg its significant enhancement in the definition of harm, extending to economic loss arising from personal or property damage; and, in the event of impairment of the (wider) environment, costs both of reinstatement and representing loss of income deriving from a direct economic interest in any use or enjoyment of the environment (cf exclusions from the Environmental Liability and CCS directives).

contracting parties.¹⁷⁸ Making allowances therefore for uncertain (but potentially vast) exposures to liability, and the consequent problems in arriving at risk evaluations, these limited regimes reflect an awkward balancing as between (especially post-Chernobyl) concerns at the potential scale of harm – still by no means necessarily covered – and the fundamental problem of securing adequate levels of financial security, particularly through private insurance.¹⁷⁹

A second, more apposite example of a broadly comparable measure, depending on how long-term risks are viewed, may be that contained in a feature of the UK's contaminated land regime, which applies liability for remediation,¹⁸⁰ expressly extending to past contamination otherwise lying beyond the reach of modern pollution control regimes. The legislation acknowledges that there may no longer be an 'appropriate person' available for service of a remediation notice.¹⁸¹ To counter the high mortality of corporate entities over time, in certain circumstances liability may attach to a successor company, though here a narrow judicial interpretation has had the effect of absolving a successor from statutory liability following privatisation.¹⁸² Moreover, a still more marked contrast between the regimes may be drawn, for, in relation to contaminated land, even in the face of the above problems of attribution no liability is placed upon state agencies,¹⁸³ and under UK law the authority has a power (rather than a duty) to carry out remediation works and recover reasonable costs.¹⁸⁴ This raises questions as to that regime's general effectiveness,¹⁸⁵ although with resource pressures on regulators, the costs burden can be partially shifted from the public to the private domain under the more negotiated processes of development planning, thereby avoiding any more coercive remediation regime.¹⁸⁶

Liability in CCS post-operational phases, by analogy with the above, is limited by both institutional and temporal constraints. The issue of survival of the operator as a legal entity is especially problematic, and here the CCS directive appears to have made a virtue of necessity, with state long-term responsibility arguably representing a notion of the public good; though it might seem that at the heart of the transfer solution lie public-private partnership arrangements that privatise potential profits whilst socialising attendant longer-term risks. To some degree, exposure from the authority's perspective will be mitigated by a statutory financial mechanism, compensating for monitoring costs over a 30-year period, with potential for a further discretionary figure to cover costs to ensure permanent containment.¹⁸⁷ Even here, such a mechanism can only deliver what the financial and

178 To a maximum of €1500 million, respectively made up by an operator obligation of €700 million (non-discretionary, the state to indemnify if insurance is not available to satisfy claims up to that figure) and total contracting party obligations up to €800 million; see M Hintergger, 'The Paris Convention 2004: A New Liability System for Europe' [2004] 3 *Environmental Liability* 116.

179 See generally J C Dow, *Nuclear Energy and Insurance* (Witherby 1989)

180 Environmental Protection Act 1990, pt IIA.

181 *Ibid* ss 78H, 78N: a further caveat, not directly analogous here, is where such person would suffer hardship if required to contribute to remediation.

182 *R (Application of National Grid, formerly Transco) v Environment Agency* [2008] Env LR 4.

183 The original proposal for the Environmental Liability Directive would have placed obligations on member states as a last resort.

184 Environmental Protection Act 1990, ss 78N, 78P.

185 See B Pontin and C Willmore, 'Displacing Remedies from Environmental to Planning Law: the Enforcement of Contaminated Land Regulation in Britain' (2006) 6 *Yearbook of European Environmental Law* 97.

186 S Vaughan, 'The Contaminated Land Regime: Still Suitable for Use?' [2010] *JPL* 142, 149–53.

187 *Dir* 2009/31/EC (n 24) art 18.1; these are barely estimable, though the European Commission may adopt indicative guidelines under art 8.2.

insurance markets are willing to bear,¹⁸⁸ and there is a regrettable lack of further explicit linkage to a more extensive establishment of trust funds from revenues generated during operational phases.¹⁸⁹ The inadequacy of private cover would be a considerable brake on the potential effectiveness of any of the above arrangements, and so in respect of long-term risk CCS depends ultimately on government willingness to act as guarantor.

6 Conclusions

Accordingly, there is a now widespread acceptance of a need to pursue the potential of CCS as an integral part of the search for solutions to the challenge of cutting carbon emissions. At the heart of societal expectations of CCS, however, lies a fundamental dilemma: as between the logical inferences to be drawn from climate science (especially concerning the imperative of reducing fossil fuel use and, in particular, coal) and the vulnerabilities of low-carbon strategies to the 'business as usual' effect (in the event that CCS proves unable to deliver upon expectations).¹⁹⁰ The problem of lock-in was referred to above, whereby dangerous levels of emissions continue, supported by an over-dependency upon the regulation of 'inherently anti-ecological activities',¹⁹¹ in the hope that CCS technology eventually proves adequate to the task of compensating failures in mitigating emissions.

Meaningful engagement with dilemmas over reducing carbon intensity for activities across a developed economy such as that of the UK ultimately depends on two key factors: first, reducing uncertainties affecting low-carbon investment (including reinventing the need to choose winners in a field that remains market-driven); and, secondly, adopting multilevel, pluralistic approaches to the development of policies and laws that seek to achieve the reduction of greenhouse gas emissions. The previous discussion from the perspective of governance was informed by the challenges that arise. It is now clear that any success in tackling climate change, as part of a wider pursuit of sustainability, will require a further blurring, as identified above, of distinctions between the public and the private domains. This is especially the case as ongoing efforts must be pursued within an energy context in which the state explores new roles and relationships with an industry which still must operate in accord with the dictates of the marketplace.

It has been argued that, without an appropriate and coherent regulatory framework, then, the complex challenges posed by CCS across each of its developmental, operational and post-operational phases cannot be addressed. In particular, regulatory approaches need to be adopted with a view to mitigating the uncertainties identified here. The field continues to be marked by wide-ranging technological, economic and investment uncertainties; as well as uncertainties as to the extent of respective health and environmental risks, including into the long term in the event of leakage from storage sites, with the further potential consequence of undermining emissions reduction objectives. In discussing instrumental responses it was further seen that complex tensions can arise, especially where different

188 Cf UK insurers subscribe to a pool operated under the agency of a company limited by guarantee for this purpose, Nuclear Risk Insurers Ltd, with insurers severally liable for respective shares of insured risks; see further, Tromans (n 176) 217–26.

189 A common feature adopted initially for pilot project sites in the US, typically in those states with extensive extraction histories such as North Dakota, Montana; also adopted in Wyoming, Louisiana; see J J Snyder, 'Obstacles to Regulation of Carbon Capture and Sequestration by US States: Can They Be Overcome?' in Havercroft et al (n 7) 204.

190 See eg J Hansen, *Storms of My Grandchildren: The Truth about the Coming Climate Catastrophe and Our Last Chance to Save Humanity* (Bloomsbury 2011), 193–94.

191 R White, *Crimes against Nature: Environmental Criminology and Ecological Justice* (Willan Publishing 2008) 232; see further R Field, 'Risk and Justice: Capitalist Production and the Environment' in D Farber (ed), *The Struggle for Ecological Democracy: Environmental Justice Movements in the United States* (Guilford Press, 1998).

legal regimes, with varied regulatory purposes, must accommodate one another. And yet in the wider analysis of regulatory prospects, lessons could likewise be drawn from experience in applying regulatory approaches to other environmental problems.

In light of the CCS directive alongside other interventions, it has been seen that a policy dynamic is emerging within the EU in the fields of climate change and now also energy, which by virtue of what has been termed institutional lock-in will increasingly influence member state actions, in compliance with legally binding emissions reductions commitments. There is a growing range of regulatory strategies in place, including under command-and-control and economic mechanisms (such as ETS), establishing a framework under which the development and control of carbon sequestration may take place. For the UK, with government seeking to have contracted commitments to currently two pilot projects in place by 2015, looking thence toward delivering CCS as a competitive element within the energy portfolio, emerging strategy will need to identify anticipated risks and costs scenarios as a precondition to securing realistic infrastructure investment across the sector through the 2020s and beyond.¹⁹² Should CCS become established, then from a storage perspective, it will be essential to pursue regulatory strategies that address identified uncertainties, especially in respect of the new risks that arise, such as harm from off-site migration into ambient air or onto surface land, and leakage to adjacent sub-surface areas. Resulting challenges directly engage with questions of regulatory capability and in this key respect the contribution of law continues to be fundamental.

¹⁹² See the latest report of the Committee on Climate Change, *Meeting Carbon Budgets, Fifth Progress Report to Parliament* (London: Stationery Office, June 2013), 'Commercialisation of CCS' 94–98.

The benefits of privacy-invading expression

PAUL WRAGG*

University of Leeds

1 Introduction

Although it has received less coverage than the Leveson Inquiry, the Joint Committee on Privacy and Injunctions, chaired by John Whittingdale MP (the Whittingdale Committee) has also investigated media freedom recently, focusing particularly on the impact of developments in privacy law. Consistent with its quiet performance, it may be said that the Whittingdale Committee's chief findings are relatively mundane, amounting to little more than a gloss on rubber-stamping common law developments. Whilst this may no doubt disappoint certain sections of the press, it is also disappointing from a legal commentary perspective, particularly for its treatment of public interest. For, although the Whittingdale Committee recognises the problematic nature of identifying this quality with precision, it does remarkably little to address it. Arguably, the imperative for deeper discussion has been provided by a discernible liberal shift in judicial reasoning in recent privacy cases whereby unauthorised media disclosures of often embarrassing facts about an individual have been treated as instances of strong free speech claims.

This shift in position is contrary to orthodox thinking on the value of privacy-invading expression, doctrinally and normatively, under art 10 of the European Convention on Human Rights (ECHR). Since the conventional debate frames art 10 as primarily concerned with protecting expression that advances democratic participation,¹ the claim that privacy-invading expression deserves minimal, if any, protection is difficult to resist.² Yet these recent cases suggest a developing disconnect between the orthodox view of the public interest in privacy-invading expression and the emergent jurisprudence. In philosophical

* Lecturer in Law and Academic Fellow of the Honourable Society of the Inner Temple, School of Law, University of Leeds. An earlier version of this article was shortlisted for the Society of Legal Scholars (SLS) Best Paper Prize 2012 and aspects of it were presented at the SLS Conference 2012. The author is grateful to conference participants for their comments and to Paul Bernal of the University of East Anglia for 'tweeting' the presentation to a much larger audience. Thanks also to David Campbell and Nick Taylor, University of Leeds, and the anonymous reviewer for their comments as well. The usual disclaimer applies.

1 See eg E Barendt, *Freedom of Speech* (2nd edn, OUP 2005) 155–62.

2 R Barnes, *Outrageous Invasions: Celebrities' Private Lives, Media, and the Law* (OUP 2010); H Fenwick and G Phillipson, *Media Freedom under the Human Rights Act* (OUP 2006) 779–86; J Rowbottom, 'To Rant, Vent and Converse: Protecting Low Level Digital Speech' (2012) 71(2) *CLJ* 355, 369–71; J W Devine, 'Privacy and Hypocrisy' (2011) 3(2) *Journal of Media Law* 169–77; P Wragg, 'A Freedom to Criticise? Evaluating the Public Interest in Celebrity Gossip after Mosley and Terry' (2010) 2(2) *Journal of Media Law* 295–320.

terms, the conceptualisation of free speech is not limited to the democratic process value alone. For instance, a variety of arguments have been presented that link protection of speech with enhanced human faculties.³ However, it may be said these arguments are not taken seriously in the academic literature as a means of rationalising judicial interpretation of art 10.⁴ Contrary to that commentary, the primary concern of this article is to show that this ‘benefits-to-self’ analysis squares with the broader dimensions of the public interest test observable in recent privacy cases and so has utility in determining the parameters of protection for privacy-invading expression in the UK. The article, therefore, advances a positivist claim that this form of expression is protected to the extent it benefits the individual. Naturally, the normative appeal of such a regime is a different matter and certain doctrinal issues arise as well. These are explored in the final part.

2 The Whittingdale Committee

In recognition of intense media and political debate about injunctions preventing privacy-invading expression and in light of high-profile breaches of injunctions on social networking sites and in Parliament,⁵ the Whittingdale Committee was tasked with reviewing common law developments in privacy under the Human Rights Act 1998, including their effectiveness, in order to consider whether statutory intervention might be necessary to clarify the law, and to review the effectiveness of the Press Complaints Commission’s ability to regulate media intrusions into privacy (recognising the duplication with the Leveson Inquiry into the culture, practice and ethics of the press). Of key significance for the following discussion, the Committee, in its report,⁶ recommended no change to the existing law through statute on the basis that, first, judicial developments had been legitimate and struck the appropriate balance between privacy and free speech⁷ and, secondly, in any event, it was doubted such a statute could provide a meaningful definition of ‘public interest’ since it may become outdated quickly or require satellite litigation concerning its interpretation. Moreover, they concluded that the term defied simple definition albeit they did suggest a future reformed media regulator might publish a clearer set of guidelines on its meaning.⁸ Overall, the Committee concluded that the courts were best placed to determine whether a particular publication was in the public interest or not.⁹ Thus, the Committee approved of the misuse of private information tort, developed by the House of Lords in *Campbell v MGN Ltd*,¹⁰ elaborated upon in *Re S*,¹¹ and similarly approved of by the European Court of Human Rights (ECtHR) in *MGN Ltd v UK*,¹² by which private information is protected where it discloses a reasonable expectation of privacy (the threshold test) and its comparable value is not outweighed by the public interest in publication. In *Re S*, the House of Lords held that the second stage of the *Campbell* test involved four elements: first, neither right has precedence over the other (ie a fully functional system of privacy is just as important to society as a fully functional system of freedom of expression); secondly, that

3 See below.

4 See eg Barendt (n 1) 13–18; Fenwick and Phillipson (n 2) 18–19.

5 In the aftermath of the Ryan Giggs affair, *CTB v News Group Newspapers Ltd* [2011] EWHC 1232 (QB).

6 *Report of the Joint Committee on Privacy and Injunctions* (the Report) (HL 2010–12 273, HC 1443, 27 March 2012) [32].

7 *Ibid* [32].

8 *Ibid* [50].

9 *Ibid*.

10 *Campbell v MGN Ltd* [2004] UKHL 22, (2004) 2 AC 457.

11 *In Re S (FC) (A Child)* [2004] UKHL 47.

12 *MGN Ltd v UK* (2011) 53 EHRR 5 [131].

‘where the values under the two articles are in conflict, an intense focus on the comparative importance of the specific rights being claimed in the individual case is necessary’; thirdly, that the justifications for interfering with each must be taken into account and, finally, any interference must be proportionate.¹³

Apart from its political value in response to those sections of the press claiming enhanced privacy protection is ‘undemocratic’,¹⁴ the findings of the Committee about the term ‘public interest’ are fairly lightweight and somewhat disappointing, particularly the timorous acceptance that it defies precise definition. Although this article focuses exclusively on the legal apparatus to determine the public interest in the speech at stake, there is also, of course, a public interest in maintaining privacy, which ought to be taken into account in determining the strength of the privacy claim. There may even be a public interest against maintaining privacy that is unrelated or only tangentially related to freedom of expression, say privacy-invading photography by a government agency to catch a benefits cheat.¹⁵ These broader notions of public interest at the different stages of the *Campbell* test were apparently overlooked by the Committee. Also, there is no real recognition that, whatever meaning is assigned to the term ‘public interest’ in a free speech context,¹⁶ it must be regarded as a continuum, rather than an absolute, with some expression only lightly engaging the term and other speech doing so much more fully. This criticism can also be made of the case law, see discussion below.¹⁷

Despite recognising the central importance of the public interest test in determining the weight of the free speech claim by noting, in particular,¹⁸ the Court of Appeal’s view that it is the ‘decisive factor’,¹⁹ it would be a stretch to describe the Committee’s scrutiny of the methodology employed by the judiciary to determine public interest as anything more than light touch. No specific cases are referred to and nothing is said about the principles informing the court’s approach to determining its meaning. It is difficult to reconcile this narrowness with the Committee’s overarching objective of determining whether the law was ‘working’²⁰ and, consequently, what is said about the term is largely superficial such that a number of points arising from the jurisprudence go unobserved. Perhaps the most significant observation missed is the tremendous scope within the current law for judicial idiosyncrasy in identifying the qualities that generate ‘public interest’.²¹ Clearly, the term is open to a broad range of interpretations, as graphically demonstrated by the diverse spectrum of viewpoints offered by those giving evidence to the Committee, particularly on the question of whether the revelation of embarrassing facts about individuals is a *prima facie* legitimate matter of public interest. Whereas several commentators (including the National Union of Journalists) suggested this was not of itself in the public interest, with one saying ‘the presumption must be that what people do in private places is their business alone, providing it is within the law’,²² others disagreed, with the most extreme viewpoint

13 *In Re S (FC) (A Child)* [2004] UKHL 47 [17].

14 See eg P Dacre, ‘The Threat to Our Press’, *Guardian* (London, 10 November 2008).

15 See eg *Wood v Commissioner of Police for the Metropolis* [2010] EMLR 1 (CA) on the limits of art 8.

16 It is recognised throughout this article that the term ‘public interest’ is a contested and controversial term in the academic literature. See eg Barendt (n 1) 27–30. For a useful discussion of the philosophical literature, see I O’Flynn, ‘Deliberating about the Public Interest’ (2010) 16(3) *Res Publica* 299.

17 S 3(b).

18 The Report (n 6) [43].

19 *ETK v News Group Newspapers Ltd* [2011] EWCA Civ 439 [23], [2011] 1 WLR 1827.

20 The Report (n 6) 7.

21 See Wragg (n 2).

22 Joint Committee on Privacy and Injunctions, ‘Oral and Written Evidence’, 16 March 2012, 30, evidence of Professor Brian Cathcart [15].

emanating (unsurprisingly, perhaps) from the Chartered Institute of Journalists who argued that the term ought to include what is of interest to the public, for no other discernible reason than the somewhat circular argument that ‘the public has demonstrated, by the tens of thousands, that it is interested in it’.²³

As this evidence demonstrates, the term ‘public interest’ is essentially a blank canvas onto which may be projected a multitude of preconceptions and predilections. It may have assisted the Whittingdale Committee to recall that, as a matter of principle, the meaning of public interest depends entirely upon the value underpinning freedom of expression since it is through that lens that the term must be seen. One of the difficulties inherent in the Committee’s conclusion that the meaning of public interest should be left to the judiciary is the discernible lack of clarity by the judiciary in specifying this value (and therefore it might have been useful for the Committee to identify it). In its assessment of freedom of speech, the Committee expresses the view that it ‘is essential for discovering new truths and thus enabling social progress; it allows for the moral and cultural self-development of individuals; and it is necessary for the flourishing of a healthy democracy’.²⁴ This view reflects the well-established philosophical appraisal of free speech²⁵ but is controversial as a description of the legal regime for determining protection since, according to the orthodox view, this plurality of free speech values does not ultimately animate judicial protection of art 10 even though references to such may be found in its rhetorical commentary on the right; instead, expression is protected to the extent it furthers democratic participation.²⁶

Thus the ECtHR jurisprudence reveals a fairly tough stance on privacy-invasive expression, as seen, for example, in the principle that speech intended only to satisfy public curiosity is not a matter of public interest regardless of how famous that individual is.²⁷ The emphasis on self-rule is also apparent in the domestic case law, for example, in Baroness Hale’s judgment in *Campbell*, such as the view that ‘top of the list is political speech’,²⁸ and the longstanding principle that the public interest is not to be confused with what interests the public.²⁹ In *Mosley*, Mr Justice Eady regarded the principle that public interest is determined by the contribution of the expression to public debate as importing ‘a very high test’.³⁰ However, in light of recent case law, the established view in the academic commentary may need modification since there appears to be a distinct softening of this stance so as to protect a broader range of privacy-invasive expression from interference, as a brief survey of the case law reveals.

In *Ferdinand v MGN Ltd*,³¹ the High Court found it was in the public interest to know that Premier League footballer and England captain Rio Ferdinand was not, as his autobiography suggested, a reformed character but had cheated on his wife prior to marriage. In *Hutcheson v News Group Newspapers Ltd*,³² public disclosure of the bare fact that the claimant had a secret second family was not only justified as part of an allegation that

23 Joint Committee on Privacy and Injunctions (n 23) 51, evidence of Chartered Institute of Journalists.

24 The Report (n 6) 9 [10].

25 See eg F Schauer, *Free Speech: A Philosophical Enquiry* (CUP 1982); K Greenawalt, ‘Free Speech Justifications’ (1989) 89 Columbia Law Review 119–55.

26 Barendt (n 1) 155; Fenwick and Phillipson (n 2) 50.

27 *Von Hannover v Germany* (2005) 40 EHRR 1.

28 *Campbell* (n 10) 499.

29 See eg *Jameel v Wall Street Journal Europe Sprl* [2007] 1 AC 359 [147].

30 *Mosley v News Group Newspapers Ltd* [2008] EWHC 1777 (QB) [131].

31 *Ferdinand v MGN Ltd* [2011] EWHC 2454 (QB).

32 *Hutcheson v News Group Newspapers Ltd* [2011] EWCA Civ 808.

he had misappropriated funds from the business he ran with his famous son-in-law, Gordon Ramsay, to support this family but also amounted to a 'strong claim' to freedom of expression because, since being dismissed from that business, Hutcheson had had a 'very public slanging match' with Ramsay, which limited his ability to complain of public discussion about him, not least because these allegations invoked the media's right to disclose matters of wrong-doing. In *Spelman v Express Newspapers Ltd*,³³ Environment Secretary Caroline Spelman MP's 17-year-old son Jonathan applied unsuccessfully for an injunction against the *Daily Star*. The nature of the private information is not specified in the judgment although it has since become known that Jonathan had been banned from playing rugby for 21 months for taking banned substances to expedite his recovery from injury.³⁴ Since he had played rugby at England under-16s level and for Harlequins, the *Daily Star* argued that there was a 'clear public interest' in reporting this matter because it highlighted 'the pressures on elite athletes from the very beginning of their sporting careers' and 'although he is only 17, he is already a role model to other youngsters who aspire to follow in his footsteps and play for their country'.³⁵ This, they said, amounted to 'a much bigger public interest issue than the fact the Claimant is the son of a Conservative politician . . . Exposing . . . pressures on those who are young at elite level informs and educates those below at grass roots level and helps to promote a culture in the public and in sport . . .', with the identity of his mother representing only an 'incidental dimension'.³⁶ The High Court does not specify the nature of the public interest in any detail in the open judgment, although it is stated that the newspaper had a 'good prospect' of establishing a public interest in the information. In *Goodwin v News Group Newspapers Ltd*,³⁷ it was a matter of public interest that infamous banker Fred Goodwin had had an affair with a colleague whilst in charge of troubled bank Royal Bank of Scotland since it raised a legitimate issue about the standard of acceptable behaviour of public figures.³⁸ Similarly, in *SKA & PLM v CRH & Persons Unknown*,³⁹ the High Court would not grant a septuagenarian businessman an injunction to prevent the defendant (an alleged blackmailer) from disclosing to the claimant's wife and grown-up children the bare fact of his adultery with, and impregnation of, the second claimant, due shortly to deliver twins. Tugendhat J concluded that it could not be said that this disclosure about the claimant (not a public figure so far as one can tell) should not be allowed because, even if the defendant was a blackmailer, it could not be said he was deprived of his art 10 rights⁴⁰ and, in this case, 'to tell a grown up child that his or her father . . . is, or is about to be, the father of twins, is speech of a high order of importance'.⁴¹ Clearly, these cases were not decided on any narrow conception of the democratic process value.

33 *Spelman v Express Newspapers Ltd* [2012] EWHC 355 (QB).

34 <www.bbc.co.uk/news/uk-17731143> accessed 16 April 2012.

35 *Spelman* (n 33) [22].

36 *Ibid* [22] and [25].

37 *Goodwin v News Group Newspapers Ltd* [2011] EWHC 1309.

38 Cf *WXY v Gewanter, Positive Profile Ltd & Burby* [2012] EWHC 496 (QB) where disclosure by the third defendant of the claimant's (a wealthy woman closely connected to a foreign head of state) alleged affair was prohibited.

39 *SKA & PLM v CRH & Persons Unknown* [2012] EWHC 766 (QB).

40 Cf Sharp J in *DFT v TFD* [2010] EWHC 2335 [23]: 'I accept Mr Tomlinson's submission that the expression rights of blackmailers are extremely weak, (if they are engaged at all)' and also *WXY v Gewanter et al* (n 38).

41 *SKA* (n 39) [77].

These may be contrasted with other recent decisions where the revelation of a sexual relationship did not amount to a discussion of public interest of itself.⁴² To the extent that the proposed publication would contain substantial details of the affair, such a finding is consistent with Lord Hoffmann's view in *Campbell* that 'the addition of salacious details or intimate photographs is disproportionate and unacceptable [and] even if accompanying a legitimate disclosure of the sexual relationship, would be too intrusive and demeaning'.⁴³ For example, in *CDE v MGN Ltd*, the first claimant, a married TV personality, had conducted what the court described as a 'quasi relationship' with the second defendant, which involved, allegedly, telephone sex and other intimate and personal discussions containing 'a good deal of flirtation and sexual innuendo' as well as the exchange of intimate photographs of themselves.⁴⁴ The *Sunday Mirror* wished to publish these details, arguing that there was a public interest in the story because the second defendant had a history of mental health problems and that the first claimant had taken advantage of her vulnerability. However, the court did not accept this argument on the facts.

Of course, the question of whether publication of the bare fact of the relationship or limited details⁴⁵ may be prevented depends not only on the asserted public interest but also on the nature of the privacy claim. The latter is invariably strengthened if both parties to the relationship wish for it to be kept secret⁴⁶ or if the claimant has young children who would be adversely affected by the revelation and/or subsequent press scrutiny.⁴⁷ Other cases show that the media may decide not to contest the non-disclosure order nor advance a public interest claim,⁴⁸ although, as the court has repeatedly said, the question of whether an anonymity or non-disclosure order should be granted is for the court to decide, not the parties, because the 'parties cannot waive the rights of the public'.⁴⁹ These cases suggest that for the public interest argument to be successful, an additional dimension beyond the mere fact of an adulterous affair is required. As the following section argues, the nature of this additional dimension may be read consistently with the 'benefits-to-self' analysis.

3 The 'benefits-to-self' argument

A) PRINCIPLE

The argument that freedom of speech deserves special protection because it benefits the self is well established in the academic literature and has been presented in a variety of forms. It is, for example, apparent in John Stuart Mill's classic argument from truth in the claim that liberty of thought and discussion 'enable average human beings to attain the mental stature which they are capable of'.⁵⁰ Likewise, pre-eminent free speech theorist,

42 *ETK* (n 19); *MNB v News Group Newspapers Ltd* [2011] EWHC 528 (QB); *CTB* (n 5); *TSE v News Group Newspapers Ltd* [2011] EWHC 1308 (QB); *MJN v News Group Newspapers Ltd* [2011] EWHC 1192 (QB); *JIH v News Group Newspapers Ltd* [2011] EWCA Civ 42; *CDE v MGN Ltd* [2010] EWHC 3308; *AMM v HXW* [2010] EWHC 2457 (QB).

43 *Campbell* (n 10) 475.

44 *CDE* (n 42) [5], [45] and [65].

45 See eg *Trimingham v Associated Newspapers Ltd* [2012] EWHC 1296 (QB) [306]–[08], in which the claimant complained unsuccessfully about various newspaper reports that she had told friends sex with Chris Huhne MP was 'wild', 'amazing', 'incredible'. Tugendhat J accepted that these descriptions were a matter of editorial judgment.

46 *ETK* (n 19).

47 *CDE* (n 42); *ETK* (n 19).

48 *MJN* (n 42).

49 *JIH* (n 42) [12].

50 J S Mill, *On Liberty in Collected Works of John Stuart Mill* vol XVIII (University of Toronto Press 1977) 243.

Thomas Emerson, specified self-fulfilment as one of four values advanced by the system of freedom of expression: 'the proper end of man is the realisation of his character and potentialities as a human being. For the achievement of this self-realisation, the mind must be free'.⁵¹ Consistent with this assessment, Raz has expressed the view that, since speech often portrays different ways of living, it benefits self-development by validating those lifestyles and, therefore, censorship equates to disapproval of those ways of life, is an insult to those who live that way and alienates them from society.⁵² Arguably, the benefits-to-self analysis is apparent in its most undiluted form in the separate works of Martin Redish⁵³ and C Edwin Baker,⁵⁴ who express the view that it is the primary – if not only – value advanced by a system of free speech. According to them, this value comprises of two elements: individual self-rule (ie expression relating to public decision-making) and self-development (ie expression relating to private decision-making), although they disagree on certain key conceptual issues. Redish, for example, conceives freedom of speech as an audience-orientated right so that speech is protected to the extent it benefits those who receive the information whereas Baker conceives of it as a speaker-orientated right⁵⁵ (although he makes a special case for the media, discussed below) and so, for example, would exclude commercial speech from protection because it is generated by market demands and not by a 'manifestation of individual liberty'.⁵⁶ Both agree that interference is justified where speech is seriously harmful.⁵⁷ Baker, for example, articulates this level of harm as excluding speech leading to violence or amounting to coercion.⁵⁸ Such a high standard of interference is important, they argue, to ensure the autonomy of the audience (in Redish's case)⁵⁹ and the speaker (in Baker's)⁶⁰ is respected. This high level of protection applies to a broad range of speech categories, not simply that defined as 'political' (as the argument from participation in a democracy tends to do)⁶¹ nor that 'objectively' defined as beneficial to the individual, since no such objective criteria could be established, certainly not by the state.⁶²

51 T Emerson, *The System of Freedom of Expression* (Vintage Books 1971) 6.

52 J Raz, 'Free Expression and Personal Identification' (1991) 11 OJLS 303.

53 See M Redish, 'The First Amendment in the Marketplace: Commercial Speech and the Values of Free Expression' (1971) 39 *George Washington Law Review* 429–73 which sets the tone for his later research, particularly 'The Value of Free Speech' (1982) 130 *University of Pennsylvania Law Review* 591–645 which presents the argument in its most complete form. See also 'First Amendment Theory and the Demise of the Commercial Speech Distinction: The Case of the Smoking Controversy' (1997) 24 *Northern Kentucky Law Review* 553–584 and 'Understanding Post's and Meiklejohn's Mistakes: the Central Role of Adversary Democracy in the Theory of Free Expression' (2009) 103 *Northwestern University Law Review* 1303–70.

54 See C E Baker, *Human Liberty and Freedom of Speech* (OUP 1989) which develops earlier themes in his work. Chapter 3 is particularly valuable in summarising his claim that the First Amendment protects the speaker's uses of speech for self-realisation and to promote change. See also 'Giving the Audience What it Wants' (1997) 58 *Ohio State Law Journal* 311–417; 'Campaign Expenditures and Free Speech' (1998) 33 *Harvard CR–CLL Law Review* 1–55; 'The First Amendment and Commercial Speech' (2009) 84 *Indiana Law Journal* 981–98; 'Is Democracy a Sound Basis for a Free Speech Principle?' (2011) 97(3) *Virginia Law Review* 515–30; 'Autonomy and Free Speech' (2011) 27 *Constitutional Commentary* 251.

55 Their disagreement unfolds in C E Baker, 'Realizing Self-Realization: Corporate Political Expenditures' and M Redish's *The Value of Free Speech* (1982) 130 *University of Pennsylvania Law Review* 646–77 and M Redish, 'Self-Realization, Democracy, and Freedom of Expression: A Reply to Professor Baker' (1982) 130 *University of Pennsylvania Law Review* 678–88.

56 Baker, *Human Liberty* (n 54) 225.

57 Redish, 'The Value of Free Speech' (n 53) 627.

58 Baker, *Human Liberty* (n 54) 59.

59 Redish, 'The Value of Free Speech' (n 53) 605–07.

60 Baker, *Human Liberty* (n 54) 47–69.

61 R Bork, 'Neutral Principles and Some First Amendment Problems' (1971) 47 *Indiana Law Journal* 1.

62 Redish, 'The Value of Free Speech' (n 53) 627–28; Baker, *Human Liberty* (n 54) 59.

This resistance to the prioritisation of ‘political’ expression reflects the belief that information relevant to private decision-making is equally important to that benefiting public decision-making⁶³ and is also due to a deep-rooted mistrust of any attempt to define ‘political expression’ reliably.⁶⁴

The appeal of the benefits-to-self argument to defending the protection of privacy-invading expression is obvious. Privacy-invading expression may be said to encourage self-reflection, personal growth and maturity in its audience, particularly where it is disapproving of celebrity excess. It may also help develop positive personality traits and generally enhance a deeper understanding of what it is to be a member of society and to be human. For those who are morally outraged, the story may become a platform by which to educate others or condemn those who engage in the problematic behaviour and, in that case, displays of media moral outrage about the personality’s behaviour reinforce their own impression of social or moral accountability. Likewise, these stories may make others feel better about themselves (say, because they have had similar experiences or feelings) because, to them, it normalises the behaviour and so develops their sense of confidence and boosts self-esteem. More mischievously, an individual may benefit from witnessing a celebrity’s downfall either because they enjoy the perverse pleasure of *Schadenfreude* or more simply because it gives them a topic of conversation by which to discover, and develop relationships with, like-minded people. These benefits are, of course, irrelevant where free speech is wholly conceived in democratic process terms but become pivotal where it is re-imagined as primarily benefiting personal development. In each case, however, it is unnecessary to precisely calculate the benefit to the individual, there only needs to be recognition of a conceivable benefit.

It is the contention of this article that these features of the benefits-to-self analysis can be located in the UK privacy jurisprudence and, therefore, it is possible to construct a positivist argument that privacy-invading expression is protected in the UK for its contribution to individual self-development. Since, at an abstract level, free speech theory should identify both the value(s) served by affording expression special treatment and the circumstances in which the state may legitimately interfere with it, these will also be set out in the following. The accounts of Baker, Redish and Emerson are informative in this respect, although they do not provide a complete solution since they adhere to certain key principles not found in the UK case law, such as the centrality of viewpoint neutrality in determining free speech protection,⁶⁵ which, as a principle, is largely alien to the UK and Strasbourg jurisprudence (see discussion below). For the following discussion, it is useful to apply an audience-orientated perspective to the analysis.⁶⁶ Such an approach is consistent with both UK⁶⁷ and Strasbourg⁶⁸ case law. For the purposes of discussion, the audience may be defined as those who want to receive the information as opposed to those who happen upon it (who may be thought of as bystanders whose interests may be largely inimical to freedom of expression).⁶⁹

63 Redish, ‘The Value of Free Speech’ (n 53) 605–07; Baker, *Human Liberty* (n 54) 33.

64 Baker, *Human Liberty* (n 54) 32–33.

65 See, for example, M Redish, ‘Commercial Speech, First Amendment Intuitionism and the Twilight Zone of Viewpoint Discrimination’ (2007) 41 *Loyola of Los Angeles Law Review* 67–132.

66 See Barendt (n 1) 23–30, for analysis of the different interest groups.

67 See eg R (*On the Application of ProLife Alliance*) v BBC [2003] UKHL 23 and discussion below.

68 See eg *Jersild v Denmark* (1995) 19 EHRR 1 [31]: ‘Not only does the press have the task of imparting . . . information and ideas: the public also has the right to receive them.’; *Krone Verlags Gmbh & Co KG v Austria* (No 3) (2006) 42 EHRR 28 [31]: ‘For the public, advertising is a means of discovering the characteristics of services and goods offered to them.’

69 Barendt (n 1) 27–30.

B) IMPACT ON THE 'PUBLIC INTEREST' TEST

By viewing the domestic privacy jurisprudence through the lens of the benefits-to-self analysis, a number of points emerge. First, although the decisions outlined in the previous section may or may not be justifiable in strict democratic process terms, it is notable that the judges in these cases do not identify the public interest in such terms and certainly not so narrowly. Instead, three distinct interpretations of the term 'public interest' are at work, all of which are notably broad: that the public has a right not to be misled and therefore the media is entitled to report *any* misleading behaviour (this right is also apparent in *Campbell v MGN Ltd*);⁷⁰ that, as role models, public figures must adhere to a high standard of behaviour; and that the media enjoys a general freedom to criticise the behaviour of others (this principle was first articulated in *Terry v Persons Unknown*).⁷¹ *Ferdinand v MGN Ltd*⁷² discloses all three interpretations, applied in varying degrees. Thus, the media's public interest claim comprised of a right to correct the public misperception created by Ferdinand's autobiography and other interviews that he was a reformed character. Of additional weight to the free speech claim was the fact that he was England captain at the time;⁷³ 'a job that carried with it an expectation of high standards',⁷⁴ and therefore spoke to the role model responsibilities that Ferdinand had, in terms reminiscent of the highly criticised Court of Appeal decision in *A v B plc* in which Lord Woolf held that 'role models' (which he described as public figures holding 'a position where higher standards can be rightly expected by the public') may be the 'legitimate subject of public attention', whether they have courted publicity or not, since 'the public have an understandable and so a legitimate interest in being told the information';⁷⁵ 'undesirable behaviour on their part can set an unfortunate example'.⁷⁶ Citing passages from *Terry*, the court in *Ferdinand* agreed that the media is entitled to criticise 'the conduct of other members of society as being socially harmful, or wrong'.⁷⁷ Clearly, the judiciary in *Ferdinand* cannot have had in mind notions of a legal wrong but, instead, some moral imperative. Yet the connection between Ferdinand's behaviour and its actual impact on the fabric of society is not established or, even, meaningfully addressed and remains entirely unspoken.

In *Hutcheson*,⁷⁸ the Court of Appeal likewise spoke of the media's 'freedom to criticise', describing it as a 'powerful' argument.⁷⁹ Applying this principle, the court held that there were two grounds on which publication was justified. First, the public nature of the dispute counted against Mr Hutcheson:

to my mind, those who choose to conduct their quarrels in such a fashion take the risk that they may not be able to insist thereafter on clear boundary lines between what is public and what is private – regardless of whether they were, hitherto, only public personalities in a very limited sense. In the present case, as

70 *Campbell* (n 10) [24].

71 *Terry (Previously LNS) v Persons Unknown* [2010] EMLR 16 [100].

72 *Ferdinand* (n 31) [65].

73 *Ibid* [89].

74 *Ibid* [89].

75 *A v B plc* [2003] QB 195 [11(xii)]; see sustained criticism in *Barendt* (n 1) 232–33; Fenwick and Phillipson (n 2) 794–800.

76 *A v B plc* (n 75) [43(vi)].

77 *Ferdinand* (n 31) [63].

78 *Hutcheson v News Group Newspapers Ltd* [2011] EWCA Civ 808.

79 *Ibid* [29].

it seems to me, there is a very real risk of a distorted and partial picture being presented to the public of this dispute⁸⁰

Thus, they concluded on this point, there was a ‘powerful interest in publication’.⁸¹ Clearly, there is a lack of sympathy for the claimant’s position but this cannot account for the obscured imperative for the audience to hold a ‘complete’ picture or, rather, why the audience would be disadvantaged by an ‘incomplete’ one. An explanation is only partially provided by the second ground, that the disclosure authenticated the allegation of wrongdoing,⁸² since this is, of itself, insufficient to wholly account for the significance of greater public knowledge about these events. Arguably, it is only by reading in the benefits-to-self analysis that the court’s overall finding of a ‘strong claim to freedom of expression in the public interest’⁸³ becomes understandable. Similarly, and as noted above, in *Spelman*, the *Daily Star* refers to the claimant as a role model when arguing that a significant public interest exists.⁸⁴ These broader notions of public interest speak to the benefit derived by the audience of knowing of wrong-doing, primarily in a moral sense, not because the knowledge impacts on public decision-making (clearly, it does not) but because it may benefit private decision-making, say, because the audience learns something valuable that they might apply to their own everyday lives: that cheating (either on a partner, in business, or in sport) is, at the very least, morally wrong and therefore should not be indulged in or, otherwise, they learn something about the individual concerned that changes their perception of them. From this, the audience might gain a deeper insight of how to behave in society and what to expect of others. Moreover, they might modify their behaviour in respect of the well-known figure publicly, for example, by not purchasing associated products or not providing support in some other way (booing at a football match, say), or, privately, in their discussions with friends about that figure.

However, the case law reveals an inconsistent approach toward the assessment of public interest. For example, the Court of Appeal decision in *ETK* discloses a fairly narrow approach to the issue. Here, the appellant and X, described as ‘figures known to the public’⁸⁵ and working in the entertainment industry, had had an affair whilst working together and the appellant ended it after his wife found out. The appellant told his employer of the affair and asked not to work with X again, which led them to dispense with X’s services, publicly announcing it was a ‘convenient moment’ to make the change.⁸⁶ At first instance, Collins J refused to grant the injunction, stating there was a public interest in the real reason for X’s departure. However, the Court of Appeal disagreed for two reasons: first, Collins J had underestimated the strength of the privacy claim (particularly by minimising the impact on the children involved) and, secondly, because there was no public interest in the disclosure, finding instead that publication would only serve ‘public prurience’.⁸⁷ Even on traditional democratic process grounds, this reasoning may be questionable given the apparent equality issue (why were X’s services dispensed with and not the appellant’s?) and, arguably, this claim is strengthened on a benefits-to-self analysis. This is not to say, though, that in the final analysis the injunction ought not to have been

80 *Hutcheson* (n 78) [45].

81 *Ibid* [45].

82 *Ibid* [46].

83 *Ibid* [48].

84 *Spelman* (n 33) [22].

85 *ETK* (n 19) [21].

86 *Ibid* [8].

87 *Ibid* [23].

granted but rather that Mr Justice Collins's initial assessment of the public interest is consistent with the benefits-to-self analysis (and, arguably, other case law).

Likewise, as noted above, the reasoning in *Mosley* discloses a narrow interpretation of the public interest, however, application of the benefits-to-self analysis would recognise the broader range of discussions that may follow from this information that might have benefited individual self-development at the time, such as, but not limited to, the appropriateness of Max Mosley's continued role as head of the Fédération Internationale de l'Automobile amidst the scandal surrounding his sex life. This would have accorded with the view of one commentator giving evidence to the Whittingdale Committee: 'Why the exposure of Max Mosley as head of F1 behaving in a quite extraordinary manner with prostitutes in a basement in Chelsea is not in the public interest . . . may not be immediately apparent to the average man in the street.'⁸⁸ In *Mosley*, Eady J concluded that the discussion did no more than satisfy public curiosity and, therefore, was contrary to the *Von Hannover* principle on public interest.⁸⁹ The benefits-to-self analysis need not conflict with this principle if 'satisfying public curiosity' is taken as a passive activity and the benefits-to-self analysis accepted as more purposeful by explicating a positive contribution to individual self-development. In this way, the *Von Hannover* principle may be respected, although the circumstances in which it may apply would be substantially reduced. It may be that, in light of the recent decisions in *Springer*⁹⁰ and *Von Hannover II*,⁹¹ this reduction in protection is in line with ECtHR case law in any event (see discussion below).

To say discussion of immoral behaviour by well-known figures is a legitimate matter of public discussion by reason of these three broader treatments of the public interest test – the public right not to be misled by the mere fact of lying or hypocrisy, the freedom to criticise doctrine and the role-model argument – is problematic to justify solely on democratic process grounds. For example, by revealing an affair, a newspaper might legitimately question the suitability of that figure for the role they hold but only if that individual holds public office.⁹² Yet, absent that aspect, the appeal to the democratic process value is less convincing. For example, the media is often expressing no more than disdain about immoral behaviour, such as alcohol abuse or adultery. Of course, this might provoke broader discussions about contemporary attitudes toward the consumption of alcohol, or its wide availability, or attitudes toward marriage or sexual promiscuity. Clearly, media discussion of these social norms is a matter of legitimate public discourse,⁹³ when expressed in general terms to identify patterns or trends, and may inform public debate on political or legal issues, such as devoting greater public money toward education on the health risks of excessive alcohol use or promiscuity, or setting minimum prices for alcohol, or harsher outcomes in divorce proceedings where there has been adultery. However, these discussions are not prevented by a ban on scrutinising the behaviour of particular celebrities.⁹⁴ Where the media's lobbying function is not engaged on the facts of a specific case, the claim to advancing democratic participation seems particularly thin. When newspapers criticise celebrity excess they are not so much concerned with what Fuller calls the morality of duty (ie those norms that identify minimum standards of behaviour in a society) but the morality of aspiration (ie those norms that articulate excellence and the

88 'Oral and Written Evidence' (n 22) 82, Alastair Brett.

89 *Von Hannover* (n 27) [65]–[66].

90 *Springer v Germany* [2012] EMLR 15.

91 *Von Hannover v Germany (No 2)* [2012] EMLR 16.

92 Eg the facts of *Trimingham* (n 45).

93 See C R Sunstein, *Democracy and the Problem of Free Speech* (Simon & Schuster 1995).

94 *Wragg* (n 2) 319.

fullest realisation of human ability and powers).⁹⁵ The state's legitimate interest in the political and legal ordering of this behaviour must end at the point where duty becomes aspiration since, as Fuller argues, it is not for the law (or politics) to determine questions of aspiration since these are more a matter of aesthetics, not least because it is impossible to say with any degree of clarity what aspirations an individual ought to have or by what standard success in this regard is to be judged; instead, the role of law is confined to the question of duty.⁹⁶ Thus, information concerning morality of aspiration has no value to democratic participation because the law cannot respond to it. Framing protection in terms of benefit-to-self renders this link nugatory.

C) WEIGHT OF THE FREE SPEECH CLAIM

Secondly, the methodology in determining the weight of the free speech claim is consistent with the view that the court cannot accurately determine the benefit to be derived by the audience. It is a discernible feature of these cases that there is a lack of judicial assessment of the comparative value of the public interest in the expression. Instead, the mere presence of a public interest is sufficient to imbue the expression with a high value. This can be seen in the *Ferdinand* and *Hutcheson* decisions, in particular, where the main thrust of the reasoning focuses on the capacity to contribute to the public interest rather than in determining the extent of that contribution for the purposes of balancing. This is arguably at odds with the *Campbell* test in two respects: first, it does not adhere to the second stage of the *Re S* elaboration of *Campbell* because there is no 'intense focus' on the actual contribution of the expression and, likewise, secondly, it conflicts with the principle that both rights are treated equally since it assigns a *potential* value to freedom of expression. The cause of this curtailment may be due to the Supreme Court's appraisal of the ECtHR's art 8/10 jurisprudence in *Re Guardian*: that, where the public interest is at stake, there is scarcely any room to interfere with freedom of expression.⁹⁷ Frankly, this is a miserly reading of the relevant case law. Although the ECtHR consistently states that the media has a duty to impart information and ideas on matters of public interest and the public has a right to receive them,⁹⁸ the analysis does not end there because the right is not absolute. On a conventional reading of the Strasbourg jurisprudence, the imperative to protect expression under art 10 does not arise simply because there is a public interest at stake (eg there may be a public interest at stake in all manner of expression that is rightly protected, such as sensitive official secrets) but where there is *sufficient* public interest at stake and, thus, the judiciary's job is not just about determining *what* the public interest is but the *weight* of it. This much is apparent from the Court of Appeal decision in *HRH Prince of Wales v Associated Newspapers Ltd*,⁹⁹ where Lord Phillips stressed that the question 'is not simply whether the information is a matter of public interest but whether, in all the circumstances, it is in the public interest that the duty of confidence should be breached'.

The decision in *Ferdinand* may be criticised for the failure to identify the public interest in the lie itself. Thus, the High Court treated the *fact* of misleading behaviour to be a matter of public interest rather than identifying what *effect* this had or might have as a matter of public interest. Moreover, the High Court appears to treat the mere fact of lying as being of not just public interest but *sufficient* public interest to outweigh the accompanying

95 L Fuller, *The Morality of Law* (Yale UP 1969) 13–15.

96 *Ibid* 15–19.

97 *Re Guardian News and Media Ltd* (2010) 2 WLR 325 [51].

98 *MGN Ltd v UK* (n 12) [141] where the court refers to *Observer v UK* (1992) 14 EHRR 153 [59]; *Thorgeirson v Iceland* (1992) 14 EHRR 843 [63]; *Bladet Tromso v Norway* (2000) 29 EHRR 125 [62]; *Jersild v Denmark* (n 68) [31].

99 *HRH Prince of Wales v Associated Newspapers Ltd* [2008] Ch 57, 124–25 [68].

invasion of privacy. Yet the fact of Ferdinand being England captain is hardly a relevant consideration in this regard and in democratic process terms is particularly problematic given the court's failure to identify the *public* nature of this role. Since the Football Association is not a public body,¹⁰⁰ nor is it publicly funded,¹⁰¹ it cannot be said that public money was at stake. Instead, the court seems to have confused what is of interest to the public (the popularity of the England football team) with what is in the public interest. It is interesting that *Ferdinand* is one of the few recent privacy decisions that the Whittingdale Committee refers to, yet these distinctive features in its reasoning go entirely unobserved despite the Committee's repeated insistence that the public interest does not equate to what interests the public.¹⁰²

Interestingly, theoretical models advocating the benefits-to-self argument do not tend to consider balancing to be a satisfactory method of determining privacy claims,¹⁰³ primarily due to the view that ad hoc balancing is an unprincipled method of deciding cases. Emerson is particularly critical of such an approach.¹⁰⁴ Instead, he conceives freedom of expression and privacy as two separate zones of activity whose boundaries abut without intersecting. For him, therefore, the judicial task is one of determining where the appropriate boundaries lie. Although the UK system is conceived differently, something of this zonal thinking could be read into the approach in recent privacy cases. It might be said that, rather than balancing the two rights, expression displaying the 'public interest' quality falls within the zone of protection irrespective of any corollary interference with private life or, even, a discernible primary motivation to titillate its audience but is outside the zone where it causes serious harm to the individual involved either because it amounts to coercion (such as blackmail)¹⁰⁵ or a threat of violence or because it touches on what Emerson calls 'the inner core of intimacy'.¹⁰⁶ Examples of the latter might be seen in the facts of *CDE* (discussed above), the humiliation experienced by Max Mosley or, more recently, in *Contostavlos v Mendabun*¹⁰⁷ where pop star and *X Factor* judge Tulisa Contostavlos obtained a non-disclosure order in interim proceedings concerning graphic footage of her engaged in a sex act, which had appeared on the internet. No public interest defence was advanced in this case and surely any attempt to construct one based on individual self-development would be bound to fail. However, media discussion of the subject matter in *Contostavlos*, that she had been videoed performing a sex act, could be justified by reference to the benefit-to-self argument, for example, because others in a similar situation might empathise with her or have subsequent views on her suitability as an *X Factor* judge (a show popular with children), as a result,¹⁰⁸ whereas it would not be on strict democratic process grounds. Thus, conceived in this way, the judicial task is not one of balancing the comparative value of the competing claims but of discerning the qualities of the expression so as to consign it to a particular zone.

100 Nor even a quasi-public body subject to judicial review on Datafin grounds (*R v Panel on Takeovers and Mergers ex parte Datafin plc* [1987] QB 815), see *R v Football Association ex parte Football League* [1993] 2 All ER 833.

101 Although it does receive public funding for community projects.

102 The Report (n 6) [79].

103 Cf Redish, 'The Value of Free Speech' (n 53) 623–25, who considers balancing is acceptable so long as there is 'a thumb on the scales' in favour of speech.

104 Emerson (n 51) 548–57.

105 *TFD v DFT* [2010] EWHC 2335 (QB).

106 Emerson (n 51) 557.

107 *Contostavlos v Mendabun* [2012] EWHC 850 (QB); see, similarly, *AMP v Persons Unknown* [2011] EWHC 3454 (TCC).

108 See also *Theakston v MGN Ltd* [2002] EWHC 137.

D) DEFINING 'PUBLIC FIGURES'

Thirdly, the sheer breadth of the term 'public figure', apparent in these decisions, is consistent with the benefits-to-self analysis since the audience may derive information crucial to their self-development just as much from knowing about the behaviour of celebrities as that of politicians or other public office-holders.¹⁰⁹ As the court recognised in *Spelman*¹¹⁰ (and as earlier recognised by the Court of Appeal in *A v B plc*),¹¹¹ the Council of Europe defines public figures as 'persons holding public office and/or using public resources and, more broadly speaking, all those who play a role in public life, whether in politics, the economy, the arts, the social sphere, sport or in any other domain'.¹¹² Although the court appears to find the definition helpful, it is hard to see what assistance the definition provides; indeed, it is hard to conceive of it as a definition at all given its all-encompassing nature. Surveying the case law, two observations arise on the characteristics required to be classified as a public figure: first, there is no requirement that the individual is in public office or directly relevant to any sort of public decision-making; second, the individual has, for whatever reason, piqued the media's interest. For that reason Samantha Brick¹¹³ is just as much a public figure as the Queen. The need for such a far-reaching definition of public figure is not obviously supported by the conventional claim that art 10 reserves special protection to speech advancing the democratic process value. The suggestion in a recent case that it is only a limited class of public figures – those holding 'important roles in national affairs' – who should expect a reduced level of privacy protection is somewhat undermined by the adjoining inclusion of footballers (and, presumably, notorious bankers and estranged fathers-in-law to celebrity chefs) within that class.¹¹⁴

E) IMPACT ON NON-DISCLOSURE ORDERS

A fourth issue arises in the context of applications for non-disclosure orders prior to full trial. Here, the court is required to apply the test set out in s 12 of the Human Rights Act 1998, of whether the claimant can demonstrate that the claim is 'more likely than not' to succeed at full trial.¹¹⁵ In such cases, however, there is a discernible trend by which the court weighs the determinable parameters of the privacy claim against an actual or hypothetical public interest claim, to be made by either the actual defendant or a hypothetical one, with the alarming result that the court treats the two parties very differently. For example, in *Terry v Persons Unknown*,¹¹⁶ Tugendhat J was particularly critical of the claimant's procedural failure to provide testimony to the prospective impact of disclosure on his private life and did not speculate on what that impact might be.¹¹⁷ By contrast, the judgment discloses a vociferous examination of the various hypothetical claims about the public interest in Terry's extra-marital affair.¹¹⁸ In cases where no defendant is present perhaps it is understandable why the court would proceed so cautiously, although, of course, even if the court issued an injunction for want of an opposing free speech claim, any interested media

109 Redish, 'The Value of Free Speech' (n 53) 644.

110 *Spelman* (n 33) [49].

111 *A v B plc* (n 75) [11 (xi)].

112 Resolution 1165 (1998) of the Parliamentary Assembly of the Council of Europe on the Right to Privacy.

113 S Brick, "There are Downsides to Looking this Pretty": Why Women Hate Me for Being Beautiful', *Daily Mail* (London, 3 April 2012).

114 *RocknRoll v News Group Newspapers Ltd* [2013] EWHC 24 (Ch) [15].

115 *Cream Holdings Ltd v Banerjee* [2004] UKHL 44 [22].

116 *Terry* (n 71).

117 *Ibid* [31]–[36].

118 *Ibid* [97]–[105].

party could later apply for the order to be varied or discharged and, in doing so, specify the nature of the public interest claim in detail.¹¹⁹ Perhaps more problematic, though, is the danger of optimisation of this hypothetical claim whereby the public interest claim is expressed in the strongest possible potential terms without any apparent discounting for the possibility that such a strong public interest claim might not be realisable on the specific facts. This process of optimisation is evident in *SKA*, for example, in the view that the information at stake was ‘certainly not trivial’.¹²⁰ In democratic process terms, this view is hard to accept but becomes more understandable when applying the benefits-to-self analysis given the claimant’s children’s obvious interest in knowing the true identity of their father. Likewise the process of optimisation seems more justifiable since it recognises the narrow circumstances in which such expression ought to be interfered with and the court’s inability to judge what is of benefit.

4 Problems

A) NORMATIVE CONCERNS

Separate to the positivist claim is the normative issue of whether privacy-invading expression *should* be protected on such grounds and the doctrinal issue of whether such protection is compatible with both the domestic and Strasbourg art 10 jurisprudence. Given that the first issue is largely open-ended in nature, the following discussion is geared more toward highlighting prospective issues and possible responses rather than anything more definitive, particularly since it is recognised that the question of whether art 10 ought to be understood so broadly deserves much more debate than it currently receives.

In the academic commentary, the benefits-to-self argument is commonly regarded as a problematic justification for freedom of speech. Schauer, for example, is particularly critical: the argument ‘suffers from a failure to distinguish intellectual self-fulfilment from other wants and needs, and thus fails to support a distinct principle of free speech’¹²¹ and that ‘even if communication is a sufficient condition for intellectual self-fulfilment, it does not follow that it is a necessary condition’: expression is not more important to understanding than experience, say, by travelling the world or driving a fast car.¹²² This is a common complaint¹²³ and closely allied to it is the problem of distinguishing ‘genuine’ free speech claims from others based on self-development, such as claims to education or housing.¹²⁴ Moreover, although it might be accepted that judges cannot decide whether particular speech improves an individual or not, it is at least arguable that certain types of speech do stunt or damage an individual’s personal and moral growth, such as hate speech or pornography.¹²⁵ In the context of privacy-invading expression, Fenwick and Phillipson suggest that the justification ‘far from inevitably opposing the right to privacy, must support it to some extent since . . . a reasonable degree of privacy is a *requirement* for individual self-development, particularly the ability to form intimate relationships, without which the capacity for individual growth would be severely curtailed’.¹²⁶ Therefore, applying Raz’s argument that speech validates different lifestyles, they argue that the restriction of privacy-

119 ‘Practice Guidance: Interim Non-Disclosure Orders’, *White Book 2012* (Sweet & Maxwell 2012) B13–074.

120 *SKA* (n 39) [77].

121 Schauer (n 25) 56.

122 *Ibid* 57.

123 See Greenawalt (n 25) 143–45; Barendt (n 1) 13–15.

124 Barendt (n 1) 14.

125 See *ibid* (n 1) 31–34.

126 Fenwick and Phillipson (n 2) 686.

invading expression 'far from suggesting condemnation or contempt for the lifestyle revealed, in fact displays respect for the ability of the individual to decide for himself whether he wishes to share his life-decisions with the public at large'.¹²⁷

These criticisms demand careful consideration although perhaps some might be more easily dealt with than others. For example, Schauer's normative concern that the justification would protect a greater range of activity than expression alone is countered by the positivist observation that art 10 is limited to protecting expression only.¹²⁸ In a privacy context, the normative issue of defining expression is unlikely to arise. Similarly, difficulties in distinguishing between a genuine free speech claim and other claims based on self-fulfilment are unlikely to trouble the courts in this field. Fenwick and Phillipson's point is perhaps most problematic to deal with since they are surely right that the public interest in individual self-development ought to be as equally significant in calculating the privacy claim as it is in determining the free speech claim. Arguably, there is little evidence in the jurisprudence that the court does sufficiently consider the public interest in privacy in determining the strength of each claimant's claim. The focus of this article has been squarely on the calculation of the public interest in freedom of expression, to the exclusion of any discussion of how the corresponding privacy claim is calculated, yet nothing in this discussion is intended to denigrate the significance of that corresponding public interest in privacy and, if the judiciary were to explicitly articulate the value of privacy-invading expression in benefits-to-self terms then it should equally articulate the value of privacy in such terms as well. Applying the logic of the argument above, the court's task is a matter of line-drawing rather than balancing. In doing so, it may be possible to map out a zone of protection that is broader than the 'zone of intimacy' to capture information that is not especially damaging to the individual concerned but nevertheless is something that the public ought not to know about. This may include, for example, knowledge that the individual is having marital or financial difficulties or has an eating disorder or is pregnant or suffers from low moods. This additional category may suit the contours of the cases outlined above which have tended to relate to expression that criticises the claimant's character in some way since it would tend to exclude behaviour that lacks that quality of 'blameworthiness' and, therefore, acts as a check on the tremendous breadth that the benefits-to-self argument would otherwise exhibit. Similarly, it has been argued that the potential adverse effect caused by reputational damage on the ability to form and/or sustain relationships with others is recognised within the domestic judicial interpretation of art 8.¹²⁹

Yet it may be said that the explicit application of the benefits-to-self argument, as a means of determining the value of privacy-invading expression and in demarcating the zone of protection, inevitably weakens the right to privacy and is inconsistent with the principle in *Campbell* and *Re S* that neither right has precedence over the other. This criticism has some merit but then it may be said, in response, that this development is already present in the jurisprudence as shown by the cases outlined above. It may also be said that the treatment of each right as presumptively equal has always had more rhetorical force than anything else and that the more recent case law suggests a revised reading of the principle is required such that it says no more than that, although free speech is the stronger right, it does not automatically succeed. More tentatively, perhaps this greater protection for freedom of expression might be defended on the basis that although the benefits-to-self argument informs the value of both privacy and freedom of expression to society, that

127 Fenwick and Phillipson (n 2) 686–87.

128 See Baker, 'Autonomy and Free Speech' (n 54) 256–57.

129 See discussion in A Mullis and A Scott, 'The Swing of the Pendulum: Reputation, Expression and the Re-centring of English Libel Law' (2012) 63(1) NILQ 27, 41–43.

value is more hypothetical in the case of privacy than freedom of expression in the context of privacy-invading expression. Although every member of public is both a consumer and potential source of news, the chances of the latter happening are considerably less than the former such that it may be said that society's greater interest is in consumption and, therefore, this should be reflected in the court's treatment of the benefits-to-self argument.

Perhaps the most pressing concern raised by explicit protection on benefits-to-self grounds is the prospect of universal application to all art 10 claims. The orthodox approach to determining the parameters of a free speech guarantee is to identify the value or values that animate the protection of *all* categories of expression, not just one. However, such an approach is problematic since it may extend protection to a range of expression that currently does not benefit from such. For example, although the claim of pornography and hate speech to furthering democratic participation is particularly weak,¹³⁰ the claim may be stronger on benefits-to-self analysis, although suppression may still be justified on account of harm. Judicial assessment of the value of public protest and unpopular political expression¹³¹ (currently precarious where expression 'offends others')¹³² would also be strengthened applying the benefits-to-self analysis. Yet, it is questionable whether there is scope to broaden protection for these forms of expression because the UK court has consistently stated it will 'keep pace' with developments at Strasbourg level¹³³ and the ECtHR jurisprudence discloses no imperative for improved protection.¹³⁴ Assuming the judiciary's attitude to these forms remains static, the preferential treatment of privacy-invading expression is not easily justified. It is surely not enough to say, on positivist grounds, that the distinction is justifiable because such discrimination is evident within the case law. However, a few points might be made in favour of such preferential treatment.

First, although theorists would object to such treatment due to the centrality of content neutrality in free speech theory, such arguments gain limited purchase when applied to the UK. Underpinning the orthodox theoretical approach to interpreting the free speech guarantee is the view that courts should adopt a position of content neutrality when determining the legitimacy of restrictions on expression. This view is a prominent feature of First Amendment analysis¹³⁵ and any failure by theory to sufficiently adhere to the doctrine provides strong grounds for criticism.¹³⁶ Since protection should not discriminate amongst content of expression it is a common feature in modern theoretical commentary that the identified value(s) apply equally to all forms of expression. However, the UK system of free speech clearly departs from the principle of content neutrality, for example, there are statutes that specifically prohibit discriminatory viewpoints.¹³⁷ Similarly, the ECtHR allows for restriction of particular viewpoints that undermine the democratic

130 See eg *Hammond v DPP* [2004] EWHC 69 (Admin); *Norwood v DPP* [2003] EWHC 1564 on hate speech; and *Belfast City Council v Miss Behavin' Ltd* (2007) UKHL 19 on pornography.

131 See eg *Abdul v DPP* [2011] EWHC 247 (Admin) on public protest and *ProLife* (n 67) on unpopular expression.

132 A Khan, 'A "right not to be offended" under article 10(2) ECHR? Concerns in the construction of the "rights of others"' (2012) EHRLR 191–204; D Mead, *The New Law of Peaceful Protest: Rights and Regulation in the Human Rights Act Era* (Hart Publishing 2010).

133 *R (Ullab) v Special Adjudicator* [2004] 2 AC 323 [20].

134 Hate speech is proscribed by virtue of art 17, *Refah Partisi v Turkey* (2002) 35 EHRR 3 [43], and member states are afforded a wide margin of appreciation where morality or public safety are at stake, *Otto-Preminger Institute v Austria* [1995] 19 EHRR 34.

135 See C Sunstein, *Democracy and the Problem of Free Speech* (Simon & Schuster 1995) 5–16.

136 See eg C E Baker 'Is Democracy a Sound Basis for a Free Speech Principle?' (2011) 97 Virginia Law Review 515, 528–29.

137 *Reg Race Relations Act 1976*, *Racial and Religious Hatred Act 2006*.

underpinning of the Convention rights.¹³⁸ There can be no real objection to a failure to adhere to content neutrality by preferring a particular type of expression.

Secondly, it might be argued that the importance of media freedom justifies such preferential treatment on the basis that the media is particularly effective in influencing audience perceptions and decision-making. This phenomenon seems to be commonly accepted¹³⁹ and has been judicially recognised, most recently by the House of Lords in the context of the broadcast media, and with some concern.¹⁴⁰ This may be contrasted with the judiciary's more dismissive approach to the importance of individual expression on influencing public attitudes, apparent in cases such as *ProLife* and *Connolly*¹⁴¹ where the respective interferences with freedom of expression were justified because of the limited impact of the expression in effecting change.¹⁴² The greater efficiency of the media in disseminating information to a broader spectrum of the populace might also underpin this differentiated approach¹⁴³ as might the orthodox view that this source of expression is also important due to its independence from government interference¹⁴⁴ or the view that the press is a 'high-level' source of authoritative information because of the considerable resources and research that (generally) underpins it.¹⁴⁵ There is some support for the special treatment of media freedom in the academic literature. Baker,¹⁴⁶ amongst others,¹⁴⁷ has argued that the separate reference to freedom of the press in the First Amendment justifies an independent meaning for the term and that the jurisprudence is 'incoherent' without such an interpretation.¹⁴⁸ However, Baker's 'fourth estate' argument differs significantly from his argument for individual freedom and is based on an instrumentalist interpretation that values the press as a checking function on government power.¹⁴⁹ There has been less appetite for privileging media freedom in the UK commentary. Barendt, for example, argues that the media's more substantial role in disseminating information is a 'thin basis' for special treatment (if 'media' is read as the traditional press) given the broad range of information providers in the market, particularly on the internet.¹⁵⁰ Elaborating on Barendt's analysis,¹⁵¹ Fenwick and Phillipson express concern that the privileging of media freedom may cut 'sharply across the other underlying rationales for speech, such as self-

138 Eg *Refah Partisi* (n 134) [43].

139 In the context of contempt of court, Barendt (n 1) 323 notes: 'the burden of proof is surely on sceptics to show that media publicity is unlikely to have any, or any significant, impact on readers, listeners, and viewers'. See also P Wragg, 'Time to End the Tyranny: Leveson and the Failure of the Fourth Estate' (2013) 18(1) *Communications and the Law* 12.

140 R (*Animal Defenders International*) v *Secretary of State for Culture, Media and Sport* [2008] UKHL 15 [29]–[30] (Lord Bingham).

141 *Connolly v DPP* (2007) EWHC 237.

142 See *ProLife* (n 67) [68] (Lord Hoffmann) and *Connolly* (n 141) [31].

143 See discussion of this model of interpreting press freedom in Barendt (n 1) 419–24.

144 See eg Schauer (n 25).

145 See recent discussion of high/low level speech in Rowbottom (n 2).

146 Baker, *Human Liberty*, (n 54) 225–49 (elaborating on the normative arguments for a 'fourth estate' reading of the First Amendment); C E Baker 'The Independent Significance of the Press Clause under Existing Law' (2007) 35 *Hofstra Law Review* 955 (where, amongst other things, he identifies themes in the jurisprudence that seem to support an independent reading).

147 P Stewart, 'Or of the Press' (1975) 26 *Hastings Law Journal* 631; F Abrams, 'The Press is Different: Reflections on Justice Stewart and the Autonomous Press' (1979) 7 *Hofstra Law Review* 563; D A Anderson, 'Freedom of the Press in Wartime' (2006) 77 *University of Colorado Law Review* 49.

148 Baker, 'Autonomy and Free Speech' (n 54) 275.

149 See Baker, *Human Liberty* (n 54) 229–34.

150 Barendt (n 1) 421.

151 *Ibid* 419–24.

development and autonomy',¹⁵² which is a particular concern in the context of privacy-invading expression.

Doctrinal support for the privileging of media protection is also doubtful. Although there is some evidence of a differentiated approach to art 10 in the Strasbourg jurisprudence, in particular, where the court has emphasised the importance of the media's role as 'public watchdog' to justify rigorous scrutiny of interferences with media freedom by member states,¹⁵³ the court has not articulated a specific distinction between media and individual freedom under art 10. Indeed, the opportunity to do so arose on the facts of *Steel & Morris v UK*¹⁵⁴ following the submission by the UK government that the applicants in a defamation case (the 'McLibel' trial), who were not journalists, should not benefit from 'the high level of protection afforded to the press'.¹⁵⁵ However, the court disagreed, emphasising that:

in a democratic society even small and informal campaign groups . . . must be able to carry out their activities effectively [since] there exists a strong public interest in enabling such groups . . . to contribute to the public debate . . . on matters of general public interest.¹⁵⁶

Thirdly, and in the alternative, it may be argued that the prospect of preferential treatment for privacy-invading expression by the application of the justification is illusory. Even if the justification were broadly applied, the resultant impact on free speech jurisprudence may not be as liberalising as it might initially appear. As noted, serious harm such as coercion or threats of violence remove privacy-invading expression from the zone of protection, so hate speech or pornography may be excluded on that basis (although the connection between pornography and serious harm is often disputed).¹⁵⁷ Indeed, this analysis may be applied to decisions involving public protest, unpopular expression and defamation. Although there are cases in which the courts have attributed a low value to the contribution of the expression to the democratic process, the decision not to uphold the art 10 claim in those cases may also be explained by the high level of actual or potential harm of the speech.¹⁵⁸

B) DOCTRINAL CONCERNS

As noted, it is now a well-established principle that the UK courts are required, under s 2 of the HRA 1998, to 'keep pace' with Strasbourg jurisprudence, 'no more, no less'.¹⁵⁹ On this basis, the protection of privacy-invading expression for its contribution to self-development must be consistent with Strasbourg jurisprudence. Yet, as Fenwick and Phillipson have argued, whereas the 'rhetorical attachment to free speech is always strong, it is by now a familiar commentary . . . to observe that clearly political speech, which more directly engages the self-government rationale, receives a much more robust degree of protection than other types of expression'¹⁶⁰ such that 'freedom of expression is valued not really as an aspect of individual autonomy or for the contribution it makes to the

152 Fenwick and Phillipson (n 2) 22.

153 See eg *Jersild v Denmark* (n 68).

154 *Steel & Morris v UK* [2005] 41 EHRR 22.

155 *Ibid* [89].

156 *Ibid* [89].

157 Emerson, for example, has previously warned of the 'serious lack of evidence' to justify the US judiciary's approach to obscenity (n 51) 467.

158 Eg *Connolly* (n 141), *Abdul* (n 131), *Hammond* (n 130), etc.

159 *Ullab* (n 133) [20].

160 Fenwick and Phillipson (n 2) 50.

flourishing of individuals, but for the part it plays in maintaining a democratic society'.¹⁶¹ For example, in *Handyside v UK*, the ECtHR said 'freedom of expression constitutes one of the essential foundations of [a democratic society], one of the basic conditions for its progress and for the development of every man'.¹⁶² This last clause has also appeared instead as 'and for each individual's self-fulfilment',¹⁶³ although without accompanying explanation of the variation.

It is not clear from the jurisprudence whether 'development of every man' or 'individual's self-fulfilment' is to be read strictly in the context of democratic participation – that is, expression which improves an individual's ability to participate in a democracy, or separately – that is, expression which improves an individual regardless of whether it improves such participation or not. Given the ECtHR's favouring of political speech it is at least arguable that the narrower interpretation is more persuasive. In the UK jurisprudence, a similar interpretation might be applied to Lord Steyn's finding in *Ex parte Simms* that freedom of expression 'promotes the self-fulfilment of individuals *in society*'.¹⁶⁴ This analysis can also be seen in Baroness Hale's treatment of free speech in *Campbell*:

intellectual and educational speech and expression are also important in a democracy, not least because they enable the development of individuals' potential to play a full part in society and in our democratic life. Artistic speech and expression is important for similar reasons, in fostering both individual originality and creativity and the free-thinking and dynamic society we so much value . . . [T]he intellectual, artistic or personal development of individuals, are not obviously assisted by poring over the intimate details of a fashion model's private life¹⁶⁵

These findings do not prevent the benefits-to-self analysis being applied overtly to justify the protection of privacy-invading expression, although they might provide some doubt that the references to self-fulfilment in the jurisprudence provide reliable precedent to justify the treatment. However, two recent ECtHR decisions, *Springer*¹⁶⁶ and *Von Hannover II*,¹⁶⁷ might lend support to this broader reading of the privacy jurisprudence.

Springer concerned the publication of several articles about X's arrest for possession of cocaine at the Munich beer festival, Oktoberfest. The ECtHR dismissed X's complaint that the articles had disproportionately infringed his art 8 rights. The reporting of crime is not problematic to justify in democratic process terms. However, the ECtHR found that the public interest in the expression 'increased' because X was an actor famous for playing a superintendent on a well-known TV show 'whose mission was law enforcement and crime prevention'.¹⁶⁸ *Von Hannover II* concerned photographs of Princess Caroline of Monaco on a skiing holiday at the time of her father, Prince Rainier, suffering from ill-health. The ECtHR dismissed the applicant's complaint on the basis that the prince's illness 'could be regarded as a matter of general interest' entitling the press to report 'on how the Prince's children reconciled their obligations of family solidarity with the legitimate needs of their

¹⁶¹ Fenwick and Phillipson (n 2) 71.

¹⁶² *Handyside v UK* (1979–80) 1 EHRR 737 [49].

¹⁶³ *Lingens v Austria* (1986) 8 EHRR 407 [41]. The *Lingens* variation has been referred to in *Tammer v Estonia* (2003) 37 EHRR 43 [59] and *Nilsen and Johnsen v Norway* (2000) 30 EHRR 878 [43].

¹⁶⁴ *R v Secretary of State for the Home Department ex parte Simms* (2000) 2 AC 115, 126 (emphasis added).

¹⁶⁵ *Campbell* (n 10) [148]–[49].

¹⁶⁶ *Springer* (n 90).

¹⁶⁷ *Von Hannover II* (n 91).

¹⁶⁸ *Springer* (n 90) [99].

private life, among which was the desire to go on holiday'.¹⁶⁹ The broader aspects of the reasoning in both cases seems more consistent with the benefits-to-self analysis than the democratic process value. In *Springer* the audience derives a benefit from discussion of the irony in X's arrest whereas the subject matter in *Von Hannover* speaks to the audience's emotional development by observing how others handle grief.

5 Conclusion

Arguably, the task of determining the meaning of public interest – the vital component in the misuse of private information tort – has proven more problematic for the judiciary than the Whittingdale Committee (and, perhaps, even the judiciary itself) give credit. In particular, the compatibility of the judiciary's approach with the orthodox view that protection under art 10 depends upon the contribution of the expression to democratic participation deserved greater discussion by the Committee than it received. It has been argued above that the current judicial treatment of the qualities peculiar to the term 'public interest' in privacy cases is contrary to the orthodox view of art 10. Though the judiciary might not view its reasoning in such terms, there is an apparently more liberal approach emerging so that speech is valued for its benefit to individual self-development.

Yet protection on benefits-to-self grounds is anathema to contemporary legal commentary. It is commonly argued that this justification is unprincipled and hugely problematic. The question of whether privacy-invading expression ought to be protected on this basis is, therefore, a separate and vexed issue. Certainly, it has not been the object of this article to suggest privacy-invading expression is important to protect (as the case law and Leveson Inquiry attests, the behaviour of the media in obtaining such stories is often unseemly and morally repugnant) but rather that the explicit application of the benefits-to-self analysis to achieve such protection has appeal for its greater consistency with established philosophical thinking on the value of free speech: it makes for a more convincing and coherent reason to protect such expression than protection on democratic process grounds. Given the criticisms in the academic commentary of the application of the argument from participation in a democracy to privacy-invading expression, it may be that this analysis represents a more intellectually honest means of justifying its protection. Privacy-invading expression may be said to benefit its audience's personal, social and moral outlook, although it is not for the court to determine the extent of this contribution (since it is plainly incapable of doing so) but rather to recognise its potential to do so. Applying the benefits-to-self analysis, it is arguable that privacy-invading expression should only be interfered with in narrow circumstances where the expression is seriously harmful, such that it amounts to violence or coercion (such as blackmail), or because it touches the 'inner core of intimacy'. This trend is also apparent in recent case law.

Perhaps the most obvious benefit derived by privacy-invading expression, which is not even discussed above, is the financial benefit to newspapers in celebrity gossip to ensure their survival. As Baroness Hale said in *Campbell*: 'It may be said the newspapers should be allowed considerable latitude in their intrusions into private grief so that they can maintain circulation and the rest of us can then continue to enjoy the variety of . . . media which are available in this country'.¹⁷⁰ Although the Whittingdale Committee considered whether the commercial viability of the press ought to play a more explicit role in the common law, it

169 *Von Hannover II* (n 91) [117].

170 *Campbell* (n 10) [143].

concluded that commercial viability, of itself, did not justify privacy-invading expression.¹⁷¹ There is much to commend this conclusion for, although commercial viability might explain the newspaper's motivation for publishing celebrity gossip, the claim to freedom of expression must ultimately hang on the value of the expression specifically and not the value of the speaker in a global sense. As one commentator noted, 'the press would also be more "commercially viable" if (to give an extreme example) it had to pay no corporation tax, or could renege on an unprofitable contractual obligation'.¹⁷²

171 In *De Geïllustreerde Pers NV v The Netherlands* [1978] ECC 164 [88], the European Commission on Human Rights stated: 'the protection of the commercial interests of particular newspapers or groups of newspapers is not as such contemplated by the terms of Article 10'.

172 The Report (n 6) [84].

Women's work: human rights, gender and social class in Hungary at the turn of the twentieth century

ISTVÁN POGÁNY*

School of Law, University of Warwick



Szerafin Thinagl (bottom left) and the graduation class in pharmacy, Ferencz József University, Kolozsvár (Cluj), Austria-Hungary, 1903

* I am grateful to the British Academy for a Small Grant in the Humanities, which enabled me to make several study visits to the Maramureş County Branch of the Romanian National Archives in Baia Mare and to the Hungarian National Library in Budapest. I should also like to thank Ms Klara Guseth, Director of the Romanian National Archives in Baia Mare, for her helpfulness and hospitality. I am indebted to Mr Béla Balogh, formerly of the National Archives in Baia Mare and to Mr Sándor Vánk, a local historian, for sharing with me their inexhaustible knowledge of Maramureş and of the materials in the archives. Mr George Barabas of Toronto, Canada, very kindly sent me photos and other materials relating to his grandmother, Szerafin Thinagl, who was the first woman to qualify as a pharmacist in the Hungarian sector of the Dual Monarchy in 1903. Translations from Hungarian are mine unless indicated to the contrary.

1 Introduction

In 'Firona', a short story originally published in the Hungarian literary journal *Nyugat*, in 1910, the eponymous heroine falls pregnant while working as a maid in a small provincial town.¹ The scattered references in the story to chestnut trees, hills and mines suggest that the town in which the story unfolds is Nagybánya where the story's author, Józsi Jenő Tersánszky, spent the early years of his life. When Tersánszky published 'Firona', Nagybánya lay in the Austro-Hungarian empire. With the redrawing of borders after the First World War, Nagybánya together with a large tranche of territory was ceded to Romania.² Today the town, which has become the administrative and legal centre of Maramureş county, is generally known by its Romanian name, Baia Mare.

In Tersánszky's story, Firona cannot remain in domestic service once her pregnancy is discovered. Although scolded by her mother and railing against Pitu, the lover who abandoned her, we are told that Firona 'didn't suffer more shame and bitterness than is customary amongst girls who share her fate'.³ The jaunty tone of this remark suggests that pregnancies out of wedlock were neither especially uncommon nor a source of everlasting disgrace, particularly amongst the poorer classes.

In Tersánszky's starkly unsentimental story, which won him immediate fame, Firona gives birth to a baby boy. Leaving the child in the care of her mother, Firona finds work as a *napszámos*, or casual labourer, on building sites. However, Firona is visibly coarsened by her new life:⁴

The dirty, heavy work of a *napszámos* greatly altered the woman. In the mornings she went off to the building sites looking slovenly and dishevelled. Later, she took every kind of ragged, sclerotic *napszámos* to be her lover. In the evenings she went drinking with them in the shops that sold *pálinka* and two or three of them would accompany her, singing all the while, on the dark footpaths beneath the vines.

The availability of work on building sites comes to an end with the heavy, persistent rains of autumn. Following the death of her mother, Firona and her little boy are left to fend for themselves in the old woman's house. Mother and child are saved from destitution by a bow-legged miner who marries Firona. Although he is kind and good-humoured, Firona comes to loathe the misshapen miner for whom, unlike Pitu her feckless ex-lover, she can never feel romantic love.

As Tersánszky's story vividly illustrates, women from the poorer classes faced a limited range of occupational choices in the Hungarian sector of the Austro-Hungarian monarchy. If, like Firona, a woman fell pregnant, her choices were fewer still. In effect, her only hope of economic security and social respectability lay in finding a husband.

At the turn of the twentieth century, poverty and social class were not the only factors that limited women's career choices in Hungary. Women from middle and upper-class families also faced a stifling set of legal and extra-legal restrictions. Their effect was to deny the overwhelming majority of Hungarian women the opportunity to complete a high school

1 J Tersánszky, *A kegyetlen primadonna* (Budapest: Osiris-Századvég 1995) 8.

2 L Kontler, *A History of Hungary* (Palgrave Macmillan 2002) 341–44.

3 Tersánszky (n 1) 9.

4 *Ibid* 9.

education (a prerequisite for university studies),⁵ to register for various subjects at university, or to pursue a number of careers, including the legal profession and the civil service.⁶

As noted by Deborah Simonton, the 'economic infrastructure of where women lived' also had a significant bearing on the range of available work.⁷ In relative terms, much of Hungary was economically underdeveloped. Working women outside Budapest and some of the more dynamic cities, such as Nagyvárad (today's Oradea), were mostly employed on the land or in domestic service.

As elsewhere in Europe, during this period, gender inequality also expressed itself in the political sphere. Women in Hungary – as in East Central Europe (ECE) more generally – were not permitted to vote in national elections or, except in very limited circumstances, to stand for public office.⁸ Throughout the inter-war period, women's voting rights in Hungary remained heavily circumscribed. The right of Hungary's women to vote in national elections on the same terms as men was not recognised until after the Second World War.⁹

This article explores women's work in Hungary at the turn of the twentieth century and the use of law in defining or, more properly, limiting the occupational and economic choices available to women. From 1867 until 1918 a large swathe of Central and Eastern Europe (CEE), amounting to well over three times the size of present-day Hungary, was governed from Budapest within the framework of the Austro-Hungarian empire.¹⁰ Hungarian law was applicable throughout this region.

Drawing on materials held in the Romanian National Archives, the article presents a case study of women's work in Máramarosziget, now the Romanian town of Sighetul Marmăției, in the early 1900s. Like Tersánszky's fictional heroine, most working women in Máramarosziget, or Sighet, were employed in domestic service or as *napszámos*, while smaller numbers earned a living as midwives, hawkers, seamstresses, nurses, washerwomen etc. Whatever form it took, women's work was almost invariably badly paid and of low status.

However, materials in the Romanian archives suggest that a small but significant minority of self-employed women came to exercise an important role in the commercial and economic life of Sighet by the turn of the twentieth century. In particular, women owned and managed a wide variety of successful businesses, including a brothel, a bath-house, numerous inns and an enterprise manufacturing suitcases.¹¹ Women in Sighet also

5 The first Hungarian high school for girls opened in 1896. See J Szapor, 'Sisters or Foes: The Shifting Front Lines of the Hungarian Women's Movements, 1896–1918' in S Paletschek and B Pietrow-Ennker (eds), *Women's Emancipation Movements in the Nineteenth Century* (Stanford UP 2004) 189, 197. Consequently, the pool of young women in Hungary who were sufficiently well educated to register at university or to undertake professional training was extremely limited.

6 Before the First World War, women's work in Britain and in Western Europe was also subject to numerous legal and extra-legal restrictions. See G Holloway, *Women and Work in Britain since 1840* (Taylor & Francis e-Library 2007), pts One, Two; D Simonton, *A History of European Women's Work* (Taylor & Francis e-Library 2003).

7 Simonton (n 6) 181.

8 Szapor (n 5) 198–200; J Malečková, 'The Emancipation of Women for the Benefit of the Nation: The Czech Women's Movement' in Paletschek and Pietrow-Ennker (n 5) 180–82; B Lorence-Kot and A Winiarz, 'The Polish Women's Movement to 1914' in Paletschek and Pietrow-Ennker (n 5) 206, 215.

9 See 4.§(1), Act VIII of 1945, available (in Hungarian) at <www.1000ev.hu/index.php?a=3¶m=8215>.

10 On the transformation of the Habsburg Empire into Austria-Hungary in 1867, see R Okey, *The Habsburg Monarchy c 1765–1918* (Macmillan 2001) 188–90.

11 See eg *Máramarosziget Város III oszt. Kerestadó kivetési lajstroma* (1902, 1903, 1906), Arhivele Naționale Maramureș, Primăria Orașului Sighet, Registre operative ale fondului, fond. nr. 46, inventar nr. 793, registrul nr 292/1902, 296/1903, 309/1906.

played a prominent role in the local economy as landlords, renting out houses or land.¹² To some degree, property, capital and business knowhow compensated for deficiencies in women's social status and legal rights, permitting a stratum of women to enjoy a significant measure of economic autonomy and personal freedom.

Women's growing involvement in the commercial life of Sighet, as elsewhere in Hungary, had been made possible by significant changes in Hungarian private law in the latter half of the nineteenth century. While women remained subject to legal and extra-legal disabilities in the *public* sphere,¹³ they achieved full equality with men in terms of their *private law* rights.¹⁴ Women, like men, could own and manage real and moveable property, enter into contracts and engage in commercial transactions of various kinds.

The extension of private law rights to women in Hungary in the second half of the nineteenth century should not be seen as the expression of a commitment to gender equality. Rather, it represented one element in a concerted effort to develop a modern, capitalist economy in the Hungarian sector of the Austro-Hungarian monarchy.¹⁵

Relatively little has been published, particularly in English, on women's work in Hungary or in the CEE area in the late 19th and early 20th centuries.¹⁶ This is part of a wider phenomenon. While the literature on women's history in the UK and in Western Europe has been described as large and expanding,¹⁷ 'Eastern European women's history has been little explored.'¹⁸ For example, Katalin Fenyves has asked rhetorically, 'Why a comprehensive history of Hungarian women is unwritten to date?'¹⁹

Scholars' inattention to gender issues in Hungary and in the CEE, until quite recently,²⁰ has been due to several factors. In the transition from communism there has been a tendency

12 See eg *Általános Jövedelmű Pótadó kivétési liststroma 1902 évre*, Arhivele Naționale Maramureș, Primăria Orașului Sighet, Registre operative ale fondului, fond. nr. 46, inventar nr. 793, registrul nr. 291/1902.

13 J Acsády, 'Remarks on the History of Hungarian Feminism' (1999) 26 *Hungarian Studies Review* 59, 60–62; J Albisetti, 'Portia Ante Portas: Women and the Legal Profession in Europe, ca 1870–1925' (2000) 33 *Journal of Social History* 825, 838, 844, 847, 848.

14 M Katona, *A Mai Érvényű Magyar Magánjog Vezérfonala* (Pozsony and Budapest: Stampfel Károly 1899) 11. §; E Csekő, 'A korlátozott előjog. Virilista nők Szekszárd r.t. városban a XX. század első évtizedeiben' in M Láczy (ed), *Nőke és férfiak . . . avagy a nemek története* (Nyíregyháza: Nyíregyháza Főiskola Gazdaságtudományi Kar 2003) 197, 198.

15 A Csizmadia, K Kovács and L Asztalos, *Magyar állam- és jogtörténet* (Budapest: Nemzeti Tankönyvkiadó 1998) 377.

16 See, however, Acsády (n 13). See also J Acsády, 'A varázstalanító emancipáció mítosza' (edited extract from the author's PhD dissertation submitted to Eötvös Loránd University, Budapest 2005) at <www.csepeleli.com/kotet/csepeleli60_acsady_judit.pdf>; Paletschek and Pietrow-Ennker (n 5) pt IV; M Feinberg, *Evasive Equality: Gender, Citizenship and the Limits of Democracy in Czechoslovakia, 1918–1950* (University of Pittsburgh Press 2006) especially chs 1–4; K Fenyves, 'When Sexism Meets Racism: the Numerus Clausus Law in Hungary' (2011) 4 *E-Journal of the American Hungarian Educators Association* 1; M Fidelis, 'Participation in the Creative Work of the Nation: Polish Women Intellectuals in the Cultural Construction of Female Gender Roles, 1864–1890' (2001) 13 *Journal of Women's History* 108. See also the sources cited in D Simonton, 'Introduction: Writing women in(to) Modern Europe' in D Simonton (ed), *The Routledge History of Women in Europe Since 1700* (Routledge 2006) 1, 11 fn 4.

17 J Purvis (ed), *Women's History: Britain, 1850–1945: An Introduction* (Routledge 1997) 1.

18 Fidelis (n 16) 109.

19 Fenyves (n 16) 14.

20 For English-language literature on women's rights in the CEE region in the post-communist era, see eg J Jaquette and S Wolchik (eds), *Women and Democracy: Latin America and Central and Eastern Europe* (Johns Hopkins UP 1998); *Women 2000: An Investigation into the Status of Women's Rights in Central and South-Eastern Europe and the Newly Independent States* (International Helsinki Federation for Human Rights 2000); 'Bending the Bow: Targeting Women's Human Rights and Opportunities' (Open Society Institute 2002) at <www.soros.org/initiatives/women/articles_publications/publications/bendingbow_20020801/bending_the_bow.pdf>; J Lukic, R Regulska, D Zavrsek (eds) *Women, and Citizenship in Central and Eastern Europe* (Ashgate 2006).

to subordinate feminist concerns to supposedly more urgent tasks, including economic development, democratisation and combating racism.²¹ Previously, during the socialist era, feminist issues were frequently marginalised by communist regimes that, in seeking a monopoly over political and social debate, shut down or appropriated feminist movements.²²

The comparative neglect of gender issues in the CEE region has also stemmed from the inaccessibility of key source materials.²³ During the socialist period, scholars, especially from Western countries, often encountered stiff bureaucratic hurdles in trying to gain access to official archives in the CEE area. In communist Romania, for example, the admission of a foreign researcher to the National Archives was considered a matter touching on Romania's political and security interests, requiring the express permission of the Ministry of the Interior.²⁴

A re-assessment of women's work in Hungary at the turn of the twentieth century is particularly timely. Feminist scholars have drawn attention to the *deterioration* of conditions facing women throughout the CEE region, whether in the workplace or in the political sphere. According to Judit Acsády, a sociologist at the Hungarian Academy of Sciences, the end of state socialism has precipitated a serious worsening of the position of women.²⁵ Similarly, Einhorn and Sever argue that, '[w]omen in this process of transition have undeniably fared badly':

The disproportionately high female share of unemployment, together with a widespread revival of nationalist and traditionalist ideologies, has had the effect of relegating women once more to the domestic sphere and has led to their growing desocialization. In this sphere women are subject to increasing violence and discrimination in the climate of frustration and social inequality often associated with a competitive marketplace.²⁶

Post-communist law, with its rhetorical emphasis on gendered justice, racial non-discrimination and the universal enjoyment of human rights, masks a less comfortable reality. Since 1990 the prospect of genuine equality both inside and outside the workplace has *receded* for large segments of society in the CEE area, particularly for women and for Europe's largest ethnic minority, the Roma.²⁷

The difficulties that women face are often compounded in countries with conservative religious traditions, such as Poland. Lorence-Kot and Winiarz emphasise that 'Polish society continued and continues to view women primarily as wives and mothers' and that 'whatever gains Polish women have made with respect to their personal rights and freedoms . . . they continue to be squeezed into an old social formula'.²⁸

The re-emergence of nationalism as a powerful political ideology across much of the CEE area has also reinforced patriarchal tendencies. As Katherine Verdery notes, nationalist political movements have encouraged efforts to 'reconfigure' the family, 'compelling women back into the nurturing and care-giving roles "natural" to their sex and restoring to men their

21 B Einhorn and C Sever, 'Gender and Civil Society in Central and Eastern Europe' (2003) 5 *International Feminist Journal of Politics* 164, 171.

22 Acsády, 'A varázstalanító emancipáció mítószá', n. 16 above, 11–14; Feinberg (n 16) 219–21; K Guenther, *Making Their Place: Feminism After Socialism in Eastern Germany* (Stanford: Stanford University Press 2010) 32.

23 Simonton, 'Introduction: Writing women in(to) Modern Europe' (n 16) 2.

24 Information supplied to the author by a retired senior archivist in Romania.

25 Acsády (n 16) 19–20.

26 Einhorn and Sever (n 21) 164.

27 K Verdery, *What Was Socialism, and What Comes Next?* (Princeton UP 1996) 79–82; I Pogány, 'Pariah Peoples: Roma and the Multiple Failures of Law in Central and Eastern Europe' (2012) 21(3) *Social and Legal Studies* 375.

28 Lorence-Kot and Winiarz (n 8) 206.

“natural” family authority’.²⁹ Thus, legal norms of non-discrimination and equal rights, which have been incorporated into the domestic law of post-communist states, are sharply at odds with conservative social, religious and political currents in many of these societies.

At the turn of the twentieth century, legal and social norms were much more closely aligned in Hungary than they are today. As previously indicated, Hungary’s women were formally excluded from the political sphere as well as from studying various subjects at university. These legal prohibitions reflected entrenched beliefs within Hungarian society concerning the appropriate role of women. Nevertheless, the liberalisation of women’s private law rights in Hungary, in the latter half of the nineteenth century, permitted a class of women to emerge who were able to exercise a substantial degree of personal autonomy and to play a significant role in the economic life of the country.

2 Law and women’s work in Hungary at the beginning of the twentieth century

(A) THE EXCLUSIONARY ROLES OF LAW, ‘BIOLOGY’ AND RELIGION

At the turn of the twentieth century in Hungary, as in much of Europe, law and bureaucratic practice served as instruments for the virtual exclusion of women from the political sphere and from various professions, as well as for the partial exclusion of women from higher education.³⁰ In effect, women faced a range of legal disabilities simply on account of their gender.

To a large extent, these disabilities reflected widely held assumptions about the unequal distribution of intellectual and other qualities between the sexes, as well as entrenched beliefs concerning the ‘natural’ role of women.³¹ For example, Géza Somogyi, a Protestant clergyman writing in Hungary, in the late 1870s, warned that emancipation and higher levels of education were twin evils that could distract women from their primary function of raising children.³² Somogyi also argued that women’s physiological characteristics demonstrated that they were intended for a life of domesticity.³³ Similarly, in a pamphlet published in Budapest, in 1902, a Catholic Bishop, Béla Mayer, asserted that university education conflicted with God’s preordained responsibilities for women and that educated women were ‘unnatural’.³⁴

Various professional associations also voiced concern at the prospect of women studying medicine, pharmacy etc.³⁵ For example, the Association of Hungarian Pharmacists warned that, if women were permitted to become pharmacists: ‘[m]any [male] customers would choose a particular pharmacy solely because of the physical allure of the pharmacist, reducing the pharmacy to the level of a coffee house or a *patisserie*’.³⁶

Despite such opposition, Hungary’s Minister of Education issued a decree in 1895 authorising universities to admit female students to degree courses in the faculties of medicine, pharmacy and the arts.³⁷ In the following years, women were also enrolled for courses in subjects including mechanical engineering, chemical engineering and

29 Verdery (n 27) 80.

30 Simonton (n 6) 241–45; R Vinen, *A History in Fragments: Europe in the Twentieth Century* (Abacus 2002) 139, 143.

31 Acsády (n 16) 4.

32 Ibid 5.

33 Ibid.

34 Ibid 5–6.

35 M Péter, *Az erdélyi gyógyszerészet magyar vonatkozásai* (Cluj: Az Erdélyi Múzeum-Egyesület kiadása 2002) 284.

36 Ibid.

37 Fenyves (n 16) 5.

architecture.³⁸ However, as discussed below, women were not permitted to register for degrees in law until 1946.

Notwithstanding passage of the 1895 decree, there remained serious obstacles to women's access to higher education in Hungary. As indicated above, there were very limited opportunities for women to obtain a high school certificate, a prerequisite for enrolling at a university. Hungary's first gymnasium for girls did not open until 1896, a year *after* women had gained the right to study at university.³⁹

Academically ambitious young women, unable to enroll at a gymnasium, could only pursue university studies if their parents could afford to hire private tutors to prepare them for the school-leaving certificate.⁴⁰ A few dozen privately educated young women took the school leaving certificate each year.⁴¹

The bulk of young women in Hungary, like their male counterparts, completed just four years of compulsory education at an *elemi* or primary school before commencing work.⁴² However, a growing proportion of girls went from the *elemi* to a *polgári*, or middle school, for an additional four years of study.⁴³ Girls who continued their education beyond this point mostly enrolled at teacher training colleges.⁴⁴

In addition, although the 1895 decree formally permitted women to study in certain university faculties, resistance within the liberal professions and in the universities was intense, resulting in shifting bureaucratic restrictions, including quotas.⁴⁵ In August 1919, the Faculty Committee of Budapest's Medical School tried to *reduce* the number of female students who were admitted,⁴⁶ while only four female students were enrolled at Budapest's University of Technology between the academic years 1920/1921 and 1929/1930.⁴⁷

The partial or total exclusion of women from certain university faculties, until 1946, had a massive impact on the scope of women's work in Hungary. In particular, the formal bar on women registering as law students at Hungary's universities, *reaffirmed* in a decree in 1927, had far-reaching consequences.⁴⁸ In pre-Second World War Hungary, a law degree was not only the route into the legal profession, it was also a prerequisite for entry to the civil service.⁴⁹ Academically able women, who might otherwise have opted to become lawyers or civil servants, were obliged to choose less prestigious and less well-paid jobs as schoolteachers, librarians or clerical workers.

In 1946, as part of wide-ranging social and economic reforms, Hungary's Parliament passed an Act that swept away virtually all of the remaining restrictions on women's access to higher education.⁵⁰ In accordance with the 1946 Act on the Admission of Women to

38 M Palasik, 'Women at the Budapest University of Technology and Economics' (Conference Paper, Charles University, Prague, 2003) 2–5, at <goliat.eik.bme.hu/~hronszky/pmea.doc>.

39 T Dobszay, 'Magyarország Kulturális Élete a Dualizmus Idején' in A Gergely (ed), *Magyarország Története a 19. Században* (Budapest: Osiris Kiadó 2003) 460, 472.

40 Szapor (n 5) 195.

41 Fenyves (n 16) 5.

42 Dobszay (n 39) 466.

43 Ibid 472.

44 Ibid.

45 Fenyves (n 16) 5.

46 Ibid 6.

47 Palasik (n 38) 7.

48 Fenyves (n 16) 12.

49 Ibid 5.

50 Act XXII of 1946 is available (in Hungarian) at <www.1000ev.hu/index.php?a=3¶m=8240>.

Universities and Colleges, women were to be admitted to *any* university or institution of higher education, other than to schools of divinity.⁵¹ The Act also stipulated that admission requirements for women applicants should be identical to those for men.⁵²

In 1948, the Hungarian Parliament passed a law prohibiting discrimination against women seeking employment in the public sector.⁵³ In addition, the statute prohibited all forms of discrimination that prevented women from obtaining the qualifications necessary to join any profession or to pursue any occupation.⁵⁴ However, limited exceptions were made for restrictions on female employment that could be justified by reference to the protection of women's health or that were based on women's 'physical capacity'.⁵⁵ Religious institutions were also exempted from compliance with the statute.⁵⁶

With the passage of Act XXII of 1946 and Act XLIII of 1948 women gained virtual equality with men in terms of their freedom to study any subject at university or to choose any occupation or profession, subject to having the appropriate academic qualifications. However, these freedoms did *not* translate into gender equality, whether in the workplace or in the political sphere. Katherine Verdery notes that, despite the egalitarian rhetoric of the communist regimes that governed the CEE region for most of the post-war period until 1989, 'the structure of power and the larger division of labor in the socialist family remained decidedly gendered'.⁵⁷ As Verdery emphasises:⁵⁸

... the state apparatus was heavily masculine. The core sectors of socialism – the bureaucracy itself, heavy industry, the army, and the apparatus of repression – were almost wholly male, especially at the apex . . . In the state bureaucracy, women overwhelmingly held clerical and secretarial functions . . . Women were indeed brought into political office, but generally at lower levels and in areas deemed appropriately female: education, health care and culture.

In its affirmation of gender equality, socialist law neither reflected nor greatly influenced the gendered character of CEE societies. This disparity between legal prescription and social and economic reality became *more* not less pronounced with the passage of time. For example, from the mid-1950s, gender was *reintroduced* as a criterion in employment policies in Hungary. Women tractor drivers, who had been the quintessential symbol of female emancipation only a few years before, were now viewed as the victims of an overly dogmatic application of socialist ideology.⁵⁹ The numbers of women working in other traditionally masculine occupations declined sharply.⁶⁰ As Verdery notes, with reference to socialist Romania, gendered distinctions were apparent in almost every sector of the economy:⁶¹

... in Romanian industry in 1985, 42 percent of the labor force was female, but women formed 80 percent of textile workers, 50 percent of those in electronics, and 30 percent of workers in machine construction; among white-collar occupations, women formed 43 percent of persons employed in science but 65

51 Act XXII of 1946 (n 50 above) 1.§(1).

52 Ibid 1.§(2).

53 2.§(1), Act XLIII of 1948. The statute is available (in Hungarian) at <www.1000ev.hu/index.php?a=3¶m=8215<http://www.1000ev.hu/index.php?a=3¶m=8325>>.

54 Ibid 2.§(1).

55 Ibid 5.§.

56 Ibid.

57 Verdery (n 27) 66.

58 Ibid 66–67.

59 E Tóth, *Kádár Leányai: Nők a szocialista időszakban* (Budapest: Nyitott Könyvműhely 2010) 70.

60 Ibid.

61 Verdery (n 27) 67.

percent of employees in the more 'feminine' jobs in culture, education, and the arts and 75 percent of health care workers.

(B) THE INCLUSIONARY ROLE OF LAW

As indicated above, public law was heavily gendered in Hungary, as in Europe more generally, at the beginning of the twentieth century. By contrast, Hungarian private law extended full equality to women.⁶² Crucially, women's economic equality had ceased to be contingent on their marital status. Thus, *A Guide to Hungarian Private Law in Force Today*, published in 1899, stated:⁶³

The financial independence of women is a fundamental principle of marriage law and flows from the freedom of the individual. A woman retains control over her personal property after entering into marriage and continues to manage her own property unless she chooses to entrust this responsibility to her husband. In the latter case, a woman may decide [at any time] to resume the management of her own property without having to offer any reasons.

In the event of divorce or of the dissolution of a marriage, a woman was entitled to the return of any property that she had brought with her into the marriage as well as of any property that she had acquired independently of her husband during the course of their married life.⁶⁴ Profits accruing from a wife's assets belonged to her alone.

The law relating to dowries also bolstered the financial independence of women whose marriages had broken down. Although the provision of dowries was discretionary, any dowry received by a husband had to be returned in full to the person who had given it in the event of a marriage coming to an end.⁶⁵ Frequently, parents, in providing their daughter with a dowry, would stipulate that the money was part of her inheritance.⁶⁶ Thus, in the event of the marriage breaking down, the dowry automatically became the property of the wife.

The process of extending full private law rights to women in Hungary had begun with the adoption of Austria's *Allgemeines Bürgerliches Gesetzbuch*, or Civil Code, in 1853.⁶⁷ The process culminated with the passage of Act XXIII of 1874 on the Age of Majority of Women.⁶⁸ This stipulated that, in Hungary, unmarried women attained adulthood at the age of 24 and that, 'all rights associated with adulthood shall enter into force' from this point.⁶⁹ Consequently, 'adulthood' rather than gender became determinative of women's private law rights. Married women were deemed to have attained adulthood on marriage irrespective of their age.⁷⁰

With the passage of the 1874 Act women acquired the same private law rights and capacities as men in the territories administered by Hungary. Thus, women, like men, 'enjoyed full legal capacity, could manage their property, enter into contracts and participate in economic life'.⁷¹ These measures preceded, by several years, the passage of the 1882 Married Women's Property Act in Britain. The 1882 statute finally invested married women in England and Wales with the same rights and capacities as their peers in

62 Katona (n 14) 11.§; Csekő (n 14) 198.

63 Katona (n 14) 11.§.

64 Ibid 100.§.

65 Ibid 98.§.

66 Ibid.

67 Csekő (n 14) 198; Csizmadia et al (n 15) 379–82.

68 For the text of Act XXIII of 1874 (in Hungarian), see <www.1000ev.hu/index.php?a=3¶m=5635>.

69 Ibid 1.§.

70 Ibid 2.§.

71 Csekő (n 14) 198.

Austria-Hungary, that is, to buy, own and sell property and to pursue legal claims independently of their husbands.⁷²

The following case study of women's work in Sighetul Marmației suggests that a significant minority of women were able to take full advantage of these new economic freedoms by the early years of the twentieth century. Such women played an active and increasingly important role in the commercial life of towns such as Sighet. By contrast, the overwhelming majority of working women in Sighet, as in other towns in Hungary, lacked capital and commercial knowhow and had received only a minimal education. Like Tersánszky's fictional heroine, Firona, such women had to accept whatever work they could find. In practice, women's occupational choices were frequently limited to domestic service, casual labouring, hawking fruit, working as washerwomen or prostitution.

3 Women's work in Sighet at the beginning of the twentieth century: a case study

(A) SIGHET AND MARAMUREȘ AT THE BEGINNING OF THE TWENTIETH CENTURY

At the beginning of the twentieth century, Sighet, known as 'Máramarossziget' in Hungarian, was the county town of Máramaros in the Hungarian sector of the Dual Monarchy.⁷³ In comparison with much of Austria-Hungary, Máramaros was poor and economically underdeveloped. A significant part of the county was mountainous or heavily forested, leaving insufficient land for the cultivation of crops. As emphasised in a report drawn up in 1900 by the *Főispán*, the most senior Crown official in Máramaros,⁷⁴ which was subsequently summarised in a local newspaper, substantial quantities of wheat, rye and barley had to be imported at considerable expense to feed the local population.⁷⁵ According to the same source, the most profitable sectors of the region's economy were the timber trade and the extraction of salt from local mines.⁷⁶ However, the salt mines and most of the timber belonged to the Crown. Consequently, the bulk of the profits from these ventures did not remain locally.

A national census conducted in 1891 found that the population of Máramaros comprised 268,281 persons of whom 122,528 were native Ruthenian speakers,⁷⁷ 33,610 native Hungarian speakers, 45,679 native German speakers, and 64,957 native Romanian speakers.⁷⁸ In reality, most of the 'German speakers' identified by the census were Orthodox or ultra-Orthodox Jews whose mother tongue was Yiddish,⁷⁹ a form of High

72 S Steinbach, *Women in England 1760–1914* (Phoenix 2005) 272–73.

73 The county is now known as Maramureș in Romanian.

74 Appointed directly by the Emperor on the recommendation of Hungary's Minister for the Interior, the *Főispán* was the government's chosen representative in each county rather than a member of the permanent civil service. See B Mezey et al, *Magyar Alkotmány Történet* (Budapest: Osiris Kiadó 2002) 348. The *Főispán* exercised far-reaching powers over every public institution in the county with the exception of the judiciary. The *Főispán*'s powers extended, inter alia, to suspending any public official whom he suspected of negligence or malfeasance. Ibid 348–49.

75 *Máramarosi Híradó*, 17 January 1900.

76 Ibid.

77 On the Ruthenian language and Ruthenian national consciousness, see P Magosci, 'The Rusyn Question' (1995) 2–3 *Political Studies* 221. See also T Kuzio, 'The Rusyn Question In Ukraine: Sorting Out Fact from Fiction' (2005) XXXII *Canadian Review of Studies in Nationalism* 1.

78 G Varády, 'Máramarosmegye' in *Az Osztrák-Magyar Monarchia Irásban és Képben* vol VI (Budapest: A Magyar Királyi Államnyomda Kiadása 1900) 437, 443.

79 Ibid.

German written in Hebrew script. Since the late eighteenth century the number of Jews in Maramureş had grown steadily, fuelled by immigration from Galicia.⁸⁰

As the census figures indicate, the two largest linguistic communities in Maramureş were the Ruthenians and the Romanians. In the course of the nineteenth century these communities, most of whose members adhered to the Greek Catholic confession, developed an increasingly strong sense of national consciousness.⁸¹ While 'nationality' and 'citizenship' are often treated as synonyms, particularly in the English-speaking world, this is far from the case in the CEE area, where *nationality* is seen as expressing an individual's sense of cultural, linguistic and political identity.

A census held in 1900 found that the population of Sighet comprised 16,901 persons.⁸² Of these, 12,658 affirmed Hungarian as their mother tongue, 2329 spoke German, 1697 Romanian and 701 Ruthene.⁸³ In addition, there were smaller numbers of Slovak, Croat and Serb native speakers, while a total of 37 persons fell into none of the above categories.

However, these statistics are at least partially misleading. The bulk of the native German speakers in Sighet were Jews whose mother tongue was Yiddish. In addition, although 12,658 residents of Sighet identified Hungarian as their first language, a significant proportion of these were Jews for whom Hungarian was their mother tongue.⁸⁴ This is clear from the fact that no fewer than 6375 of Sighet's residents – amounting to almost 38 per cent of the population – declared in the census held in 1900 that they were Jewish by religion.⁸⁵

Although Hungarian law at the turn of the twentieth century treated religious affiliation as *irrelevant* to questions of linguistic or national identity, Hungarian society was less scrupulous. Despite the passage of an emancipation law, in 1867,⁸⁶ Jews remained subject to varying degrees of discrimination unless they converted to Christianity. For example, Jews were much less likely than non-Jews to secure jobs in the state or local bureaucracy, in the judiciary, in the prosecution service or in the army, sectors that were seen as the preserve of the Hungarian middle and, especially, of the upper-middle class.⁸⁷ In these circumstances, many of Hungary's Jews continued to opt for careers in industry and commerce, where they were well established.⁸⁸

Although deeply divided on religious, political and other grounds, Jews constituted the largest single confessional group in Sighet by 1900. According to the census held in that year, 6375 persons identified themselves as 'Israelite', that is Jewish, by religion.⁸⁹

80 P Wrobel, 'The Jews of Galicia under Austrian-Polish Rule, 1867–1918' (2002) 29 *Galizien German Descendants Newsletter*, at <www.jewishgen.org/galicia/html/jews_of_galicia.pdf>.

81 I Berend, *History Derailed: Central and Eastern Europe in the Long Nineteenth Century* (Berkeley and Los Angeles: University of California Press 2003) 114–19.

82 *A Magyar korona országának 1900 évi népszámlálása*, Budapest 1902–, 304, in Arhivele Naționale Maramureş, fila nr. 304.

83 *Ibid* 305.

84 By contrast, conservative Jews, who were much more numerous in the northeast and east of 'historic' Hungary, were wary of integration with non-Jews and of surrendering their distinct culture and identity. Such Jews remained faithful to their Orthodox and ultra-Orthodox roots. See eg R Patai, *The Jews of Hungary* (Detroit: Wayne State UP 1996) 322–24.

85 *A Magyar korona országának 1900 évi népszámlálása* (n 82) 305.

86 For the text of the emancipation law see Act XVII of 1867, available (in Hungarian) at <www.1000ev.hu/index.php?a=3¶m=5318>.

87 F Fejtő, *Hongrois et Juifs* (Paris: Éditions Balland 1997) 156–58; I Pogány, 'Poets, Revolutionaries and Shoemakers: Law and the Construction of National Identity in Central Europe During the Long 19th Century' (2007) 16(1) *Social and Legal Studies* 95, 103–05.

88 Fejtő (n 87) 162–64.

89 *A Magyar korona országának 1900 évi népszámlálása* (n 82) 305. The second largest denominational group in the town comprised Greek Catholics (4586) and the third largest Roman Catholics (4314).

Given the size of Sighet's Jewish population, it follows that a large proportion of the town's working women were also Jewish. Although a minority enjoyed middle-class incomes and lifestyles, most belonged to the poorer classes, in common with the overwhelming majority of Sighet's Jewish residents. Working Jewish women in Sighet made a living, variously, as market vendors, seamstresses, midwives, domestic servants and as casual labourers or *napszámós*.⁹⁰

Most working women in Maramureş, as in Hungary more generally, worked from necessity rather than choice. As emphasised by Katalin Fenyves: 'in 1900, more than two thirds of Hungarian women over the age of 14 were not married, which most of the time meant that they had to support themselves'.⁹¹

The following analysis of women's work in Sighet is divided into two parts. In part 3(b), the article examines the range of work available to women from the poorer classes in Sighet at the beginning of the twentieth century. In part 3(c), the article considers the variety of work undertaken by middle and lower-middle class women in the town in the early 1900s.

(b) POOR WORKING WOMEN IN SIGHET AT THE BEGINNING OF THE TWENTIETH CENTURY

Drawing a line between 'poor' and 'other' working women in Sighet is a difficult and, to some degree, arbitrary exercise. There are almost infinite gradations of 'poverty' and 'wealth', while such notions are, to some degree, inherently subjective. However, at the turn of the *twentieth* century, various occupational categories were associated with modest incomes and low social status. This was true, for example, of domestic servants and casual labourers, groups who did not pay personal income tax directly.⁹² Consequently, women who earned a living as *napszámós* or as household servants are included here in the category of poor working women.

With respect to other predominantly female, or non-gendered, occupations, the threshold of 'poverty' used in this article is that of women paying *less* than 10 korona income tax annually. In 1902, by way of comparison, carpenters in Sighet were taxed, variously, at 14 to 34 Korona income tax, cabmen at 12 to 24 Korona income tax, barbers at 17.60 to 55.24 Korona income tax, butchers at 21.80 to 313.60 Korona income tax, doctors at 30 to 60 Korona income tax and midwives at 4 to 12.60 Korona income tax.⁹³ The average income tax paid by some 15 midwives working in Sighet was just 6.9 Korona.⁹⁴

90 Records of illegitimate births in Sighet, at the beginning of the twentieth century, show that a large proportion of Jewish single mothers were working as domestic servants, as *napszámós* or in some other menial capacity. In 1900, for example, 15 out of 57 mothers who gave birth to illegitimate children in Sighet were Jewish. Of these, four worked as *napszámós* (entry nos 162, 412, 602 and 654), five were employed as maids or cooks (entry nos 233, 305, 500, 508 and 569), one was a fruit seller (entry no 537) and one a seamstress (entry no 667). The remainder of the single Jewish mothers had no listed occupation (entry nos 524, 548, 584, 658). See *Magyar Anyakönyv* (Sighet, 1900), Arhivele Naționale Maramureş, Colecția Matricole Confesionale, fond nr. 19, inv. nr. 24, registrul nr. 8/1894–1923, Baia Mare - nașteri. As is clear from Table 5 below, in 1902 Jewish women also featured prominently amongst low-income, self-employed women in Sighet subject to 3rd Class Personal Income Tax.

91 Fenyves (n 16) 5.

92 In accordance with Act X of 1883, *napszámós* were exempted from paying personal income tax. See Act X of 1883, available (in Hungarian) at <www.1000ev.hu/index.php?a=3¶m=6088>. Although domestic servants were subject to income tax, this was usually paid on their behalf by their employers.

93 See, variously, entries 3–11 (carpenters), 17–58 (cabmen), 208–18 (barbers), 552–66 (butchers), 587–99 (doctors), *Máramarosziget Város III oszt. Kerestadó* (1902) (n 11).

94 *Ibid* entries 808–24.

In most cases, poor working women in Sighet had received a perfunctory education, at best. Not infrequently, these women were illiterate. Attendance at *elemi* or primary school had been compulsory in Hungary since the passage of Act XXXVIII of 1868.⁹⁵ However, as indicated above, primary school consisted of only four classes. At the age of ten many children, girls as well as boys, entered the job market.⁹⁶

In addition, levels of school attendance were strikingly uneven. For example, illiteracy amongst ethnic Romanians in Hungary, in the first decade of the twentieth century, was estimated at 50–60 per cent and amongst Ruthenes at 70 per cent.⁹⁷ Consequently, poor working women in Sighet and Maramureş, a significant proportion of whom were of Romanian or Ruthene ethnicity, often lacked even a basic education. For the most part, these women could only hope to find work as domestic servants, washerwomen or casual labourers.

(i) Registers of births in Sighet and the surrounding area

From 1895 an annual register of births, or *Anyakönyv*, was compiled by each municipality in the Hungarian portion of the Dual Monarchy.⁹⁸ The *Anyakönyv* records the name, place of birth and religion of every newborn child's parents as well as the occupation of the father. However, in cases where a child was illegitimate at birth and where the child's father declined to acknowledge paternity, the *Anyakönyv* lists the occupation of the mother. Consequently, registers of births for Sighet and for other towns in the region, including Baia Mare, can tell us much about the work that was available to single women from the poorer classes in Maramureş and in neighbouring counties at the beginning of the twentieth century.

Tables 1–3 provide details of the occupations of single mothers who gave birth in Sighet in 1898, 1899 and 1900, as listed in the town's registers of births. Table 4 lists the occupations of single mothers who gave birth in Baia Mare in 1900.

Table 1: Occupations of unmarried mothers in Sighet, 1898⁹⁹

Occupation of mother	Number of single mothers
Household servants (cooks, maids)	20
<i>Napszámos</i> or day labourers	11
Washerwomen	3
Seamstress	1
Restaurateur	1
No listed occupation	13
Total	49

95 For the text of Act XXXVIII of 1868 (in Hungarian), see <www.1000ev.hu/index.php?a=3¶m=5360>.

96 Dobszay (n 39) 466.

97 Ibid.

98 Previously, the task of maintaining records of births, marriages and deaths had been the responsibility of the various religious denominations.

99 See *Magyar Anyakönyv* (Sighet, 1898), Arhivele Naționale Maramureş, Colecția Registre Stare Civilă, fond nr. 743, inv. nr. 1371, reg. nr. 104/1898.

Table 2: Occupations of unmarried mothers in Sighet, 1899¹⁰⁰

Occupation of mother	Number of single mothers
Household servants (cooks, maids)	30
<i>Napszámos</i> or day labourers	10
Fruit hawker	1
Seamstresses	3
No listed occupation	7
Total	51

Table 3: Occupations of unmarried mothers in Sighet, 1900¹⁰¹

Occupation	Number of single mothers
Household servants (cooks, maids)	22
<i>Napszámos</i> or day labourers	18
Washerwomen	2
Hospital nurse	1
Fruit hawker	1
No listed occupation	13
Total	57

Table 4: Occupations of unmarried mothers in Baia Mare, 1900¹⁰²

Occupation	Number of single mothers
Household servants (cooks, maids)	18
<i>Napszámos</i> or day labourers	6
Washerwomen	3
Basketweaver	1
Seamstress	1
No listed occupation	2
Total	31

As is clear from Tables 1–4, most unmarried mothers were working when they fell pregnant. Almost without exception, these women had poorly paid, physically strenuous jobs that required little or no education. A significant proportion of the single mothers earned a living as *napszámos*: 22.44 per cent, 19.6 per cent, 31.57 per cent and 19.35 per cent, respectively. However, domestic servants constituted by far the largest single occupational group amongst the unmarried mothers: 40.81 per cent, 58.82 per cent, 38.59 per cent and 58.06 per cent, respectively.

The preponderance of household maids and cooks amongst the single mothers suggests a correlation between domestic service and sexual exploitation. As Viktor Karády has pointed out, domestic servants in Hungary were generally young, impressionable and unsophisticated; they often came from villages and rural settlements and lacked friends or relatives in the towns where they mostly found work. These considerations, as well as the

100 *Magyar Anyakönyv* (Sighet, 1899), Arhivele Naționale Maramureș, Colecția Registre Stare Civilă, fond nr. 743, inv. nr. 1371, reg. nr. 105/1899.

101 *Magyar Anyakönyv* (Sighet, 1900) (n 90).

102 *Magyar Anyakönyv* (Baia Mare, 1900), Arhivele Naționale Maramureș, Colecția Registre Stare Civilă - exemplarul II, fond. nr. 743, inventar nr. 1371, registrul nr. 6/1899–1900, Baia Mare – nașteri.

fact that they lodged with their employers, rendered female domestic servants peculiarly susceptible to sexual exploitation.¹⁰³

The exposure of domestic servants to sexual and other forms of abuse emerges quite clearly from statistics of suicide for the period. For example, amongst female domestic servants in Budapest in the first decade of the twentieth century, levels of suicide were more than *ten times* higher than amongst *napszámos* and almost *three times* higher than amongst factory workers.¹⁰⁴ These figures suggest that domestic servants were subject to extremely high levels of stress, harassment and social isolation.

The comparatively large numbers of female *napszámos* who fell pregnant outside marriage also calls for explanation. Female *napszámos*, like the fictional Firona, belonged to one of the very lowest strata of society and may have been less constrained by the rigid social conventions that governed female sexuality amongst the middle classes and amongst the 'respectable' poor, that is, peasants, household servants and industrial workers.

(ii) personal income tax files

At the turn of the twentieth century, officials in every town in the Hungarian portion of the Dual Monarchy maintained personal income tax files. In accordance with Act XXIX of 1875, there were *four* classes of personal income tax.¹⁰⁵ *Napszámos* or casual labourers, domestic servants, factory workers, shop assistants, office clerks and various other poorly paid occupational groups were subject to 1st Class Personal Income Tax. This was calculated in the form of a 'head' tax rather than as a proportion of the employee's earnings.¹⁰⁶ The tax was paid by employers, who could then deduct it from their employees' wages.¹⁰⁷

Owners of real property, including land and buildings, were subject to 2nd Class Personal Income Tax.¹⁰⁸ As discussed below, records for Sighet show that a significant number of women earned very substantial sums from renting out property. Some of these women were amongst the most affluent members of the community.

Civil servants, local government officials, clergymen, teachers and others in permanent employment and in receipt of a regular salary were subject to 4th Class Personal Income Tax, which rose to a maximum of 10 per cent on annual earnings.¹⁰⁹ However, as explained in part II(a) above, as a consequence of wide-ranging restrictions on women's education and on their entry into various sectors of employment, very few women, other than teachers, were subject to 4th Class Personal Income Tax during the period in question.¹¹⁰

Significantly, well over 100 women are listed in the 3rd Class Personal Income Tax files compiled in Sighet in the early 1900s. Persons subject to 3rd Class Personal Income Tax ranged from tinsmiths, musicians, market stall-holders, midwives and rag and bone men, at one end of the socio-economic spectrum, to lawyers, doctors, pharmacists and successful

103 V Karády, *Zsidóság és társadalmi egyenlőtlenségek, 1867–1945* (Budapest: Replika Kör 2000) 151–53.

104 Karády (n 103) 154–55.

105 See Act XXIX of 1875, available (in Hungarian) at <www.1000ev.hu/index.php?a=3¶m=5684>.

106 *Ibid* 2, 7§§.

107 *Ibid* 11§. In 1883, as noted previously, *napszámos* were exempted from this tax. See Act X of 1883 (n 92).

108 See 2, 7§§, Act XXIX of 1875 (n 105).

109 *Ibid* 2, 26§§.

110 In any event, the 4th Class Personal Income Tax files for Sighet, for the period under consideration, are missing from the National Archives in Baia Mare. The files may have been lost in transit from Sighet to Baia Mare, in the course of transferring the county archives, or they may have been destroyed in Sighet during or shortly after the Second World War. According to a former archivist, Soviet troops, who were stationed in Sighet at the end of the war, tore pages indiscriminately from the archives to use as cigarette paper, to light fires and for other purposes.

traders at the other. The tax was levied on self-employed persons other than those subject to 1st or 2nd Class Personal Income Tax. The 3rd Class Personal Income Tax files for Sighet convey a sense of the variety of occupations that were available or, equally importantly, *unavailable* to women in this period.

As the data in Tables 5 and 6 and in the corresponding footnotes suggests, *ethnicity* as well as *gender* played a significant role in determining the occupations available to – or favoured by – particular segments of the working population of Sighet. The large number of German (ie Jewish) names amongst the self-employed women listed in Tables 5 and 6 reflected the heavy concentration of Jews in many branches of commerce and industry in Sighet, as in Hungary more generally. As discussed above, this distinctive pattern of economic activity was the consequence of several factors. In the first place, Jews were not welcome in the army, the judiciary and the state bureaucracy. The latter, in particular, had become the preserve of the Hungarian lower gentry. In addition, despite a gradual broadening of vocational opportunities for Jews in Hungary, following passage of the 1867 emancipation law, many Jews preferred to remain in trades or occupations with which they or their families were already familiar.¹¹¹

Sighet's 3rd Class Personal Income Tax files for the years 1900 and 1901 are missing from the archives in Baia Mare, likely casualties of the chaos that descended upon the region in the latter stages of the Second World War. However, the 3rd Class Personal Income Tax file for 1902 lists 1006 persons.¹¹² Of these 102, a little more than 10 per cent, are women.¹¹³

It is clear from the file that many occupations were heavily gendered. With just a single exception, *all* of Sighet's tinsmiths, carpenters, timber merchants and cattle butchers were male.¹¹⁴ Similarly, *all* of the town's doctors, pharmacists and lawyers were men, even though women had been permitted to study medicine and pharmacy (but not law) at Hungarian universities since the mid-1890s.¹¹⁵ The situation remained unchanged two years later; there was not a single woman amongst Sighet's physicians, pharmacists and lawyers in 1904.¹¹⁶ By contrast, all of the midwives and seamstresses in Sighet – occupations that were of low status and that were badly paid – were female in 1902 and in 1904.¹¹⁷

The first female pharmacist to qualify in the Hungarian sector of the Dual Monarchy, Szerafin Thinagl, graduated from the Ferencz József University in Kolozsvár (now the Romanian city of Cluj) in 1903.¹¹⁸ The first female doctor to qualify at a Hungarian university, Dr Sarolta Steinberger, obtained her diploma in Budapest in 1900.¹¹⁹ However,

111 Pogány (n 87) 103–07.

112 *Máramarossziget Város III oszt. Kerestadó* (1902) (n 11).

113 Because of their multiple commercial interests, two of the women, Mrs Antal Szkankó and Szabina Almási, are listed twice each. See entry nos 408, 431 (Mrs Antal Szkankó) and entry nos 221, 471 (Szabina Almási) *ibid.*

114 Of 14 cattle butchers (ie *mészáros*), one was female. See *ibid* entry no 564 (Mrs Mihály Sreiner).

115 The piecemeal and controversial opening up of higher education in Hungary to women is discussed above, pt 2(A).

116 See *Máramarossziget Város III oszt. Keresetadó kivétési lajstroma* (1904), Arhivele Naționale Maramureș, Primăria Orașului Sighet, Registre operative ale fondului, fond. nr. 46, inventar nr. 793, registrul nr. 304/1904.

117 For details of the tax paid by midwives in Sighet see entry nos 808–24, *Máramarossziget Város III oszt. Keresetadó* (1902) (n 11). The midwives paid varying amounts of income tax (in Korona): 12.60, 5, 6, 12, 6, 6, 6, 6, 5, 5, 12, 4, 6, 4, 8. For details of the tax paid by seamstresses in Sighet see *ibid*, entry nos 929–33. The seamstresses paid the following amounts of income tax (in Korona): 2, 2, 6, 8, 4. Details of the income tax paid by midwives and seamstresses in Sighet, in 1904, are available at *Máramarossziget Város III oszt. Keresetadó* (1904) (n 116), entry nos 622–29 (midwives) and nos 716–18 (seamstresses).

118 Péter (n 35) 285.

119 J Honti, 'Száz éve végzett a budapesti egyetemen az első orvosnő' (15 September 2000) 1(1) *Semmelweis Egyetem újság*, at <www.ujzag.sote.hu/se200001/20000112.html>.

as indicated above, lingering prejudice against the notion of women pharmacists and doctors, a series of shifting bureaucratic obstacles and the late development of Hungarian secondary education for girls severely limited the number of women who were able to study pharmacy, medicine or other subjects at Hungary's universities, particularly before the First World War.¹²⁰

Some of the self-employed women listed in Table 6 were extremely successful. However, they represent only a small *minority* of self-employed women in Sighet. As is clear from Tables 5 and 6, most of the town's self-employed women had relatively modest incomes. Overall, working women in Sighet – comprising large numbers of domestic servants and *napszámós* as well as much smaller numbers of teachers and self-employed women paying 3rd Class Personal Income Tax – were poor, even by the standards of the day.

As indicated above, poor or low-income working women are defined in this article as women working in occupations such as domestic service or casual labouring (*napszámós*) as well as self-employed women paying *less* than 10 korona income tax annually. Wherever possible, the actual tax, that is, 'Decided Tax', paid by individual women is listed in Tables 5 and 6 rather than the 'Estimated Tax', as previously calculated by tax officials. However, in cases where no tax was eventually levied, for whatever reason,¹²¹ the Estimated Tax is given.

Table 5: Occupations of low income, self-employed women in Sighet subject to 3rd Class Personal Income Tax, 1902ⁱ

Occupation	Number of women	Decided Tax in korona
Bakers ⁱⁱ	2	6, 6
Bookbinder ⁱⁱⁱ	1	4.68
Lessees (land) ^{iv}	2	0.30, 0.10
Lessees of orchards, kitchen gardens etc ^v	12	1*, 3*, 0.40*, 1.5*, 1.5*, 0.25*, 0.30*, 5.50*, 1*, 0.30*, 0.50*, 0.1*
Market stall-holders ^{vi}	4	4, 4, 4, 4
Midwives ^{vii}	14	5, 6, 6, 6, 6, 6, 5, 5, 7*, 5.40*, 4, 6, 4, 8
Miscellaneous ie illegible ^{viii}	2	6, 6
Pork butchers (<i>hentes</i>) ^{ix}	7	6.40, 6.40, 2, 7.20, 5*, 3, 7.40*
Seamstresses ^x	5	2, 2, 6, 8, 4
Tobacco vendors ^{xi}	2	0.34, 4.86
Vendor of stewed vegetable dishes (<i>főzelék</i>) ^{xii}	1	8
Women's hairdresser ^{xiii}	1	8
Total^{xiv}	53	

*denotes Estimated Tax

120 The impediments to the education of women doctors and pharmacists at Hungary's universities, after 1895, are discussed above, pt 2(A). The outbreak of the First World War had a positive impact on attitudes towards the education of women. Under the exigencies of war, restrictions on the admission of women to study pharmacy and medicine were considerably relaxed. Péter (n 35) 287–88.

121 Such reasons could include the death of a self-employed person *before* personal income tax was levied for the year in question or the collapse of a taxpayer's anticipated profits.

Notes to Table 5

- i *III oszt. Keresetadó Kivetési lajstroma 1902 évre* (n 11).
- ii Ibid entry nos 427, 430 (Mrs Jakab Hovits, widow; Mrs Ferenc Renyák).
- iii Ibid entry no 520 (Mrs József Haan).
- iv Ibid entry nos 77, 127 (Mrs János Fekete, widow; Mrs Ferenc Zékány, widow).
- v Ibid entry nos 134, 136–37, 142, 157, 159–61, 170, 172, 178, 189 (Mrs János Bodnár, widow; Mrs József Budaházy; Mrs József Csek; Mrs Sándor Darvay; Leona Joódi; Mrs Erzsébet Lányi; Mrs Róza Lendrai; Mrs Gábor Magos, widow; Mrs János Schneider, widow; Mrs György Schroff, widow; Mrs Mihály Tamás, widow; Mrs József Pálfi, widow).
- vi Ibid entry nos 440, 444–45, 463 (Mrs Rifke Davidovits, widow; Mrs Mózes Feürvorker, widow; Mrs Mózes József Fisch; Mrs Judáh Szöbel).
- vii Ibid entry nos 809–18, 820–24 (Mrs Gál Csiszár, widow; Mrs Lipót Ferenczi; Mrs Gusztáv Ilnicki; Mrs Sándor Karantai; Mrs Péter Kelemen; Mrs Imre Kiss; Mrs Lipót Lóvinger; Mrs Ferencz Mentel; Mrs Lajos Patai, widow; Emma Todorovics; Mrs Sándor Verőczi, widow; Mrs István Veres; Francziska Weiser; Regina Viner).
- viii Ibid entry nos 506, 573 (Francziska Saüer; Mária Yakubinyi).
- ix Ibid entry nos 408–09, 411–13 (Mrs Antal Szkankó; Mrs György Szénási; Mrs Mária Pelenta; Mrs József Tóth; Mrs Miklos Antal; Teréz Hampigel).
- x Ibid entry nos 929–33 (Erzsébet Tevicki; Háni Kertenbaum; Matild Kirchmayer; Mrs Gábor Magos; Mrs Leopold Madler).
- xi Ibid entry nos 310, 312 (Mrs Áaron Szimovics; Ms Kati Wiesel).
- xii Ibid entry no 368 (Mrs Joel Lax).
- xiii Ibid entry no 340 (Ms Teréz Klein).
- xiv One of the women listed in Table 5, Mrs Antal Szkankó, is also listed in Table 6 as she had two separate sources of income, one of which was much more lucrative than the other.

**(C) MIDDLE AND UPPER INCOME WORKING WOMEN IN SIGHET AND THE SURROUNDING
AREA AT THE BEGINNING OF THE TWENTIETH CENTURY**

(i) 3rd Class Personal Income Tax

As indicated in Table 5, more than half of the self-employed women in Sighet subject to 3rd Class Personal Income Tax earned comparatively little, although some of these women may have had supplementary sources of income. By contrast, a number of self-employed women listed in Table 6 were financially successful. Of the 50 women in Table 6, 12 paid more than 40 Korona income tax in 1902.

It is clear from the figures in Table 6 that the highest-earning, self-employed woman in Sighet was Szabina Almási, who ran the town's brothel. Together with a second woman, Mrs Imre Egyed, Szabina Almási also co-owned a bath-house and bar. On the assumption that Szabina Almási had a 50 per cent share in the bath-house, she paid a total of 340 Korona personal income tax in 1902. This suggests an income that was high even by the standards of the most affluent members of the local community. By way of comparison, in the timber trade, which was one of the most lucrative businesses in Sighet, only a partnership formed by two timber merchants, Glasner and Schüler, paid more taxes than Szabina Almási.¹²² The third most successful timber merchant in Sighet, Zsigmond Klein, paid 225 Korona income tax,¹²³ while the next most successful, a partnership of two businessmen, paid 295.85 Korona income tax.¹²⁴ Sex, like timber, was a valuable commodity in Sighet at the turn of the twentieth century.

¹²² III oszt. Keresetadó Kivetési lajstroma 1902 évre (n 11) entry no 327 (Glasner and Schüler). The partnership paid 1928 korona income tax in 1902.

¹²³ Ibid entry no 332 (Zsigmond Klein).

¹²⁴ Ibid entry no 334 (Krauer and Kaufman).

The data in Table 6 and in the accompanying footnotes indicates that a substantial proportion of the better earning, self-employed women in Sighet were Jewish. For example, almost *all* of the women who owned bars in the town had German ie Jewish names.¹²⁵ The comparatively large number of Jews amongst the more successful, self-employed women in Sighet is explicable by the fact that Jews had been largely free to pursue careers in commerce and industry, in Hungary, while many other sectors of employment had been partially or totally closed to them.¹²⁶ Consequently, many Jewish women would have grown up in families where close family members (fathers, husbands etc) were shopkeepers or traders. Such women were much more likely than their non-Jewish peers to inherit a business or to decide to go into business themselves. And, having done so, they were much more likely to be successful.

As indicated in Table 6 and in the accompanying footnotes, a sizeable proportion of the higher earning, self-employed Jewish women in Sighet were widows. In many cases, these women ran businesses in which they had previously worked alongside their husbands, gaining valuable commercial experience.

By contrast, the relative under-representation of non-Jewish women amongst the more successful, self-employed females listed in Table 6 stemmed, at least in part, from the fact that commerce and industry were generally deemed unsuitable occupations by Hungary's Christian middle and, in particular, upper-middle class. As noted by François Fejtő, the Christian upper-middle class in Hungary preserved an essentially 'feudal mentality'.¹²⁷ Men from this sector of Hungarian society, particularly those who did not own landed estates that could assure them an adequate income, favoured careers in the state bureaucracy. Women from this privileged class were expected to content themselves with a purely domestic role as wives, mothers and home-makers. As emphasised by Katalin Fenyves, *most* working women in Hungary, in the early 1900s, were either unmarried or widowed. Thus, 'women often were expected to leave their jobs immediately after marriage in order to, on the one hand, fulfil their maternal function and on the other, not to take jobs from men'.¹²⁸

Table 6: Occupations of middle and upper income self-employed women in Sighet Subject to 3rd Class Personal Income Tax, 1902ⁱ

Occupation	Number of women	Decided Tax in korona
Bakers ⁱⁱ	1	16
Bar-owners ⁱⁱⁱ	16	48.97, 70.90, 30, 29.74, 10, 16, 27.89, 48.10, 50, 20, 68.41, 14.96, 27.25, 35.77, 19.16
Barber ^{iv}	1	25*
Brothel keeper ^v	1	300
Butcher (beef etc) ^{vi}	1	21.80
Cabbies (1 or 2 horses) ^{vii}	6	12, 12, 12, 12, 12, 24
Chimneysweeps ^{viii}	2	30, 15
Contractor, supplying domestic servants ^{ix}	1	15
Estate agent ^x	1	23.35

125 See n iii to Table 6 below.

126 Fejtő (n 87) 156–58, 162–64.

127 Ibid 163; Patai (n 84) 367.

128 Fenyves (n 16) 4.

Grain merchants ^{xi}	2	35, 10
Manufacturer of suitcases ^{xii}	1	21.36
Midwives ^{xiii}	3	12.60, 12, 12
Mill owners ^{xiv}	2	30, 30
Miscellaneous ie illegible ^{xv}	1	20
Owner of bath-house, with attached bar ^{xvi}	1	80
Owner of engineering firm ^{xvii}	1	100
Owners of restaurant, with attached café, bar etc ^{xviii}	2	100, 52.42
Pastry maker (mézes kalácsos) ^{xix}	1	26
Tobacco vendors etc ^{xx}	2	219.88, 42.10
Vendor of stewed vegetable dishes (<i>főzelék</i>) ^{xxi}	1	10
Wool merchant ^{xxii}	1	15
Watchmaker ^{xxiii}	1	20
Total	53	

*denotes Estimated Tax

Notes to Table 6

- i *III oszt. Keresetadó Kivetési Iajstroma 1902 évre* (n 11).
- ii Ibid entry no 431 (Mrs Antal Szankó).
- iii Ibid entry nos 470, 775, 477, 482–84, 486, 490–91, 495, 497, 502–05 (Mrs Mayer Brandstein; Mrs Jakab Gojd; Mrs Aurel Junger; Mrs Adolf Krámer, widow; Berta Kronstein; Mrs Iszák Lax, widow; Laura Lorber; Mrs Volf Róth; Róza Schnébalg and Mrs Dávid Lax; Záli Schwarcz; Czáli Szuk; Czili Weisz; Chive Weisz; Róza Weisz; Mrs Chaim Wisder).
- iv Ibid entry no 212 (Mrs Mór Lichtenberg, widow).
- v Ibid entry no 221 (Szabina Almási).
- vi Ibid entry no 564 (Mrs Mihály Sreiner).
- vii Ibid entry nos 26, 31, 39, 46, 48, 56 (Mrs Alter Hoch; Mrs Lieb Klein; Mrs Herczli Sachter, widow; Mrs József Sámuel, widow; Mrs Mayer Spiczter, widow; Mrs János Lebnicki, widow).
- viii Ibid entry nos 420, 424 (Mrs László Demkó, widow; Mrs János Zablahovsky).
- ix Ibid entry no 254 (Letti Rozner).
- x Ibid entry no 195 (Mrs Leó Altman).
- xi Ibid entry nos 544, 546 (Ilona Lieberman; Mrs Léb Steinmetz).
- xii Ibid entry no 468 (Mrs Bertalan Weisz).
- xiii Ibid entry nos 808, 811, 819 (Julia Balázs; Julia Gáspár; Mrs Gusztáv Schreiner).
- xiv Ibid entry nos 549–50 (Mrs Mihály Fodor; Mrs Lajos Sipos).
- xv Ibid entry no 574 (Anna Pakler).
- xvi Ibid entry no 471 (Mrs Imre Egyed – Szabina Almási is listed as co-owner of this establishment. However, she has already been recorded in Table 6 as the owner of the brothel).
- xvii Ibid entry no 567 (Mrs Mihály Balthazar).
- xviii Ibid entry nos 968–69 (Róza Zombory, Mrs Salamon Teüchman).
- xix Ibid entry no 571 (Mrs István Pap).
- xx Ibid entry nos 308, 309 (Ms Fani Halpert; Mrs Janos Halmik).
- xxi Ibid entry no 366 (Mrs Baruch Fisch, widow).
- xxii Ibid entry no 378 (Mrs Hankel Kaufman).
- xxiii Ibid entry no 584 (Mrs Rosenthál and partner).

(II) INCOME FROM RENTING OUT PROPERTY OR LAND

As indicated in Table 7, a number of women in Sighet earned substantial sums from renting out property or land. These women, many of whom were widows, played an important role in the local economy providing, variously, rooms, apartments, houses or land for rent. It is clear from records held in the archives that a large proportion of the population of Sighet lived in rented accommodation, often in overcrowded and squalid conditions. Frequently, an entire family occupied a single room in a boarding house.¹²⁹

In addition to renting out property or land women, such as those listed in Table 7, provided regular work for maids, housekeepers, woodcutters, gardeners and chimney sweeps, as well as for an assortment of craftsmen, including carpenters, glaziers, roofers, plasterers, plumbers etc. Women who rented out property or land constituted a significant source of employment in a region that was notably poor and underdeveloped, in comparison with much of Austria-Hungary, and in which seasonal work was commonplace. In view of the higher life expectancy of women, many of Sighet's *rentier* class were widows, as confirmed by Table 7.

The extent to which women who rented out property or land were directly involved in the day to day management of their assets cannot be determined from information in the tax files. However, it is reasonable to assume that, to varying degrees, many of the women listed in Table 7 would have taken some measure of responsibility for the running of their properties, whether by overseeing the collection of rents, the choice of tenants or the upkeep of houses, outbuildings, fences etc.

In contrast to running an inn or a shop, for example, renting out land or houses, particularly if all of the physical labour and much of the day-to-day administration were delegated to others, represented a socially acceptable form of economic activity for middle or even upper-middle class women in Hungary at the turn of the twentieth century, particularly widows or spinsters. Unlike most ordinary commercial transactions, renting out land or property was seen as an acceptable way of conserving a family's assets as well as a perfectly legitimate means of producing additional income.

Table 7: Selection of women in Sighet who derived a significant income from renting out property or land, 1902¹³⁰

Name of taxpayer	Decided Tax on earnings from renting out property (in korona)	Decided Tax on earnings from renting out land (in korona)
Klára Kahán	390.66	
Mrs Bertalan Szabó, widow	268.80	
Mrs János Papp, widow	227.58	
Mrs János Szabó, widow	231.62	99.29
Mrs Pál Timár, widow	136.20	60.30
Mrs Ferenc Kahán	273.28	4.99
Esther Asztalos	62.72	123.79
Mrs Joachim Pál, widow	1,481.53	49.71

129 For example, in 1902, a dwelling at 86 Fecske utca, Sighet, was owned by a religious studies teacher, Jenő Bányai. The building, which the author has seen on visits to Sighet, is a modest-sized villa. According to the tax files for 1902, the residents of the dwelling comprised the owner as well as the following persons: the widow Seregi (housekeeper); Mari (the maid); tenants – János Kondrai (unemployed waiter) and wife; János Komáromi (carpenter) and wife; Mór Lebovics (shoemith) and wife; Mihály Turicska (woodcutter); see *I-II oszt. Sziget Város Keresetadó Kivetési listroma 1902 évre*, fila nr. 287, inventar 793, fond nr. 46, Arhivele Naționale Maramureș, 140. The number, names or ages of any children that would also have been living in the building, together with their parents, are not given in the tax files. However, it is known that at least some of the tenants had children. For example, Mór Lebovics, who happens to be the author's great grandfather, had two daughters by the date in question.

130 *Általános Jövedelmű Pótadó kivetési listroma 1902 évre*, Arhivele Naționale Maramureș (n 12).

4 Conclusion

In Margit Kaffka's novella, *Színek és Évek* (*Colours and Years*), originally published in Budapest in 1912, the heroine, Magda, reflects on her younger self as the wife of a lawyer in a small provincial town. Magda shares a comfortable apartment with her husband and a live-in maid. However, despite her apparent good fortune and anxiety to fulfil the role of a dutiful wife, Magda finds life empty and monotonous, an endless round of tedious domestic chores:¹³¹

It's been like this for a year now, almost exactly the same every day. I get up in good time and, panting and straining, work my way through the little, three room apartment. In the sitting room, I dust all the porcelain objects on the little shelves. Then I start in the kitchen. I clean the table-lamp, polish the silver, arrange the cups and saucers, tidy everything away and sweep out the room. Before noon, I beat the carpets, polish the doorknobs, prepare vegetables for lunch and harry the maid. I instruct her, nag her and keep my eyes on her while she works. Is this how it's always going to be from now on? Why yes, for the rest of our lives.

By contrast, Jenő, Magda's husband, leaves the apartment each day, 'breakfasted, satisfied, smiling', a freshly lit cigar in his mouth, having run his eye over the newspaper.¹³²

Like most wives of that era Magda lives vicariously, through her husband. Magda's notion of fulfilment is to see Jenő achieve the success that she craves for him. For Magda, a husband's professional or material achievements are the ultimate testament to a wife's skill, as well as her deepest source of satisfaction. For example, Magda recalls how, as a young wife, she 'cried out softly, clapped her hands and threw [her] arms' around her husband's neck when he told her that he expected to be appointed *Alispán*, or sub-prefect, of the county:¹³³

I felt that all my dreams and ambitions were to be realised and that there was nothing more that I could wish for after this . . . Yes, I thought to myself, a woman can accomplish anything with a man. A woman can get everything she needs or desires through a man, provided that she instils enough confidence in him, badgers him and employs all her stubbornness and feminine wiles. After all, isn't that a woman's true role?

The notion that a woman could be professionally successful in her own right simply does not occur to Magda. Such unorthodox ideas are beyond her powers of imagining.

For most middle and upper-middle class Hungarian women, like Magda, the *emancipatory power* of Hungarian law, which had invested the country's women with full equality by 1874, was incomparably weaker than the *inhibitory power* of social convention. Married women, particularly middle and upper-middle class married women, were expected to be decorative companions, home-makers, mothers, not autonomous human beings with the same rights, freedoms and ambitions as their husbands.

For the overwhelming majority of middle-class married women in Hungary, at the turn of the twentieth century, their personal and professional lives were defined by family pressure, social convention and religious dogma rather than by abstract legal rights. *In principle*, Hungarian law of the period endowed women with wide-ranging economic and commercial freedoms as well as the possibility, however notional, of pursuing university

¹³¹ M Kaffka, *Színek és Évek* (Budapest: Kossuth Kiadó 2007) 48.

¹³² *Ibid.*

¹³³ *Ibid.* 83.

studies and of joining professions including medicine and pharmacy. *In practice*, the education available to most middle-class Hungarian women, in the early 1900s, was only sufficient to prepare them for a life of child-rearing, the efficient management of their households and, in their leisure hours, the perusal of romantic novels. At the turn of the twentieth century, the provision of secondary education for young women in Hungary remained hopelessly limited. Along with the factors mentioned above, this prevented many women from taking advantage of their legal rights and of realising their aspirations of achieving 'professional' status:¹³⁴

In 1890, the overall number of women of professional classes was 26,792, 0.3 percent of the whole female population, while the number of professional men was 155,850 in the countries of the Hungarian Crown, including Croatia and Slavonia. However, this data is misleading in as much as many of the above women of professional classes were midwives (almost ten thousand) or governesses, the latter coming mainly from abroad. A decade later, 1.5 percent of all working women had intellectual careers (33,873) . . .

For women from poorer families, particularly widows and unmarried women, the situation in Hungary, at the turn of the twentieth century, was entirely different. Economic necessity rather than boredom or the quest for intellectual fulfilment drove poorer, single women to seek work. As indicated above, the gendered nature of law and bureaucratic practice in this period severely inhibited women's access to higher education, to the professions and to careers in public service, while permitting or even encouraging poorer women, *who constituted the overwhelming majority of women in Hungary*, to take physically strenuous, tedious and badly paid jobs as domestic servants, *napszámos*, washerwomen etc. In effect, lower-class Hungarian women, like Tersánsky's Firona, had to work to support themselves until they could assure their future – and that of their children – by attracting a suitable marriage partner.

Ultimately, most middle-class women in Hungary, as elsewhere in Europe, were also dependent on finding a husband. In *Színék és Évek*, Magda's social and economic status rests almost entirely on her successful, ambitious lawyer husband. Despite the economic and social gulf that separates them, Firona and Magda are alike in their abject dependency on their respective partners.

Nevertheless, as evidence from the Romanian National Archives suggests, a minority of middle-class women in Sighet, many of whom were widows, enjoyed substantial economic autonomy and a reasonable or even handsome standard of living. These women owned and managed a range of businesses that included mills, shops, restaurants, bars, concessions, a bath-house and a brothel. As indicated above, a large proportion of these self-employed women were Jewish, a consequence of centuries of legal and de facto constraints imposed by host societies on the occupational choices available to Jewish communities in the CEE area. Contrary to the familiar anti-Semitic trope, Jews' over-representation in commerce and industry was not so much the consequence of freely determined individual choices as the legacy of centuries of systematic exclusion from most other forms of work and numerous constraints on Jewish ownership of land.¹³⁵

Although still denied the right to vote or to stand for election, at the turn of the twentieth century, a substantial number of women in Hungary achieved striking levels of

¹³⁴ Fenyves (n 16) 4.

¹³⁵ See eg P Johnson, *A History of the Jews* (Harper & Row 1988) especially 311, 319. See also, the sources cited in n 87.

personal autonomy, self-esteem and social recognition through the medium of Hungarian private law, which afforded them equal rights with men. The economic independence and commercial success of such women, most of whom were self-employed, compensated for their lack of public law rights, including the right to vote or to stand for election, and for a prevailing social climate that remained heavily gendered.

A more efficient derivative action system in China: challenges and opportunities through corporate governance theory

DR JINGCHEN ZHAO*

University of Sussex

Abstract

The derivative action is an exception to the rule in Foss v Harbottle¹ that was introduced with high expectations during a Chinese company law (CCL) reform that came into effect in 2006. It is vital to combat the abuse of derivative action in China, a country with an emerging corporate governance model, a weak legal system and inefficient enforcement measures. This article examines several deficiencies in derivative actions, with the purpose of arguing for a more effective and positive derivative action rule for the benefit of shareholders and their companies, and also for the development of CCL and the reconstruction of the Chinese corporate governance system. Corporate governance theories will be discussed as theoretical supports for a more efficient and enforceable derivative action system in China.

Key words: China; company law; corporate governance; derivative claim; shareholders.

1 Introduction

The rule in *Foss v Harbottle*² provides a negative answer to the question of whether an individual shareholder is able to bring a complaint before the court if an irregularity has been committed in the course of a company's affairs, or if some wrong has been done to the company. Derivative action is an exception to the rule in English law that defines the company itself as the proper person or legal entity to bring action when seeking redress against wrongdoers who are in control of the company. Statutory derivative action is designed to overcome inadequacies of common law in the area of shareholder remedies, including the prohibitive cost of litigation, the restrictive standing requirement, and uncertainties concerning the rule and its exceptions.³ However, it may also lead to manifest injustices when it goes wrong, not only due to the essence and logic of the law but also to the efficiency and defects of the procedure itself.⁴ Derivative action provides that if a

* Senior Lecturer in Corporate and Commercial Law, University of Sussex, Jingchen.Zhao@sussex.ac.uk. I am grateful for the comments of an anonymous referee.

1 *Foss v Harbottle* (1843) 2 Hare 461.

2 *Ibid.*

3 I M Ramsay and B B Saunders, 'Litigation by Shareholders and Directors: An Empirical Study of the Statutory Derivative Action' (Melbourne Law School Legal Studies Research Paper No 250, Centre for Corporate Law and Securities Regulation, University of Melbourne 2006) 10–14.

4 See C Hale, 'What's Right with the Rule in *Foss v Harbottle*?' (1997) 2 *Company Financial and Insolvency Law Review* 219; B R Cheffins, 'Reforming the Derivative Action: The Canadian Experience and British Prospects' (1997) 2 *Company Financial and Insolvency Law Review* 227.

shareholder can establish a case in which the action harming the company constitutes a fraud on the minority and where the wrongdoers control the company, he or she will be permitted to take derivative proceedings which are derived from the company. This trend has been followed by a number of countries in recent legislation, such as the UK,⁵ Australia,⁶ Canada,⁷ New Zealand⁸ and Singapore.⁹

In spite of the fact that China does not follow the rule of common law, after much consideration and public consultation, the derivative action was introduced with high expectations during a CCL reform in 2005, coming into effect on 1 January 2006. The rule, which was proposed to the Standing Committee of the National People's Congress in February 2005, was adopted to address one of 12 defects in the old CCLs. The changes regarding derivative action are part of a series of changes surrounding shareholder protection, especially for minority shareholders. The Chinese legal system is still distinguished from other jurisdictions by the underlying Chinese market and its legal institution in the context of a control-based corporate governance model.¹⁰ The new provisions for derivative action were embodied in both the CCL 2006 and the Chinese Securities Law 2006. The provisions showed both convergence and divergence characteristics when compared with statutory derivative action provision in other legislations. The reforms in the new CCL reflect a compromise in relation to the efficiency of CCL and procedure law in terms of the protection of shareholders, especially minority shareholders.

It is vital to combat the abuse of derivative action in China, a country with an emerging corporate governance model, a weak legal system and inefficient enforcement measures. This article examines several deficiencies in derivative action, with the purpose of arguing for a more effective and positive derivative action rule for the benefit of shareholders and their companies, and also for the development of CCL and the reconstruction of the Chinese corporate governance system. Corporate governance theories, namely agency theory and path dependence theory, will be discussed as theoretical arguments for a more efficient and enforceable derivative action system in China.

2 Emerging derivative actions

The rule in *Foss v Harbottle* is that, in general, 'where a wrong is done to a company, only the company may sue for the damages caused to it; a shareholder has no right to bring an action on behalf of the company in order to protect the value of his shares'.¹¹ Two principles stem from this general rule, namely the elementary principle¹² and the principle of majority rule.¹³ A derivative claim emanates from the 'fraud on the minority' exception to the rules enshrined in *Foss v Harbottle*, which was designed by the court to be a device to allow minority shareholders to bring a claim against those who control the company. In such a

5 See Companies Act 2006, s 236.

6 Corporations Act 2001, pt 2F.1A

7 See Canadian Business Corporations Act 1985, ss 232 and 242.

8 See Companies Act 1993 (NZ), ss 165 and 166; see also P Prince, 'Australia's Derivative Action: Using the New Zealand Experience' (2000) 18 Company and Securities Law Journal 493; S Watson, 'A Matter of Balance: The Statutory Derivative Action in New Zealand' (1998) 19 Company Lawyer 236.

9 Companies Act 1994 (Singapore), cap 50, s 216.

10 X Huang, 'Derivative Action in China: Law and Practice' (2010) 6 Cambridge Student Law Review 246, 246–47; see also J Zhao and S Wen, 'Promoting Stakeholders' Interests in the Unique Chinese Corporate Governance Model: More Socially Responsible Corporations?' (2010) 21 International Company and Commercial Law Review 373.

11 *Newbigin-by-the-Sea Gas Co. Ltd v Armstrong* (1879) 13 Ch D 310.

12 *Konamaneni v Rolls-Royce Industrial* [2002] 1 BCLC 336.

13 *Carlen v Drury* (1812) 1 Ves & Bea 149.

legal claim, the minority shareholders, instead of enforcing a right that belongs to them, are vested in and therefore derived from the company¹⁴ and are therefore in the representative capacity of the company.¹⁵ In the UK, the new statutory derivative action, as provided in ss 260–64 of the Companies Act 2006, is built on principles developed by the courts over the last 150 years. With regard to civil law countries, in Germany, for instance, AktG, the legislative framework governing derivative claims for public limited companies, was set out in s 148 of the German Aktiengesetz introduced by art 1 nr 15 of the Act on Corporate Integrity and Modernisation of the Right of Avoidance (*Gesetz zur Unternehmensintegrität und Modernisierung des Anfechtungsrechts, UMAG*).¹⁶

The emergence of derivative action in China is the result of the rapid development of the stock market. Chinese stock exchanges, including the ones in Shanghai and Shenzhen, developed very quickly after 1990 and by 2008 China's stock market had grown to become the second largest in Asia after Japan. By 1 April 2013, there were 2491 listed companies with a total of capital of 38.805 trillion renminbi (RMB) (£3.88 trillion).¹⁷ One of the interesting characteristics of the Chinese capital market is the large percentage of small investors (owning less than 1 million RMB in cash-worth of shares), accounting for 99 per cent of the total number of capital accounts.¹⁸ The average turnover rate is seven times higher than rates in more mature markets.¹⁹ In order to counter the booming capital market packed with unsophisticated individual investors, the problems of inadequate shareholder protection are becoming increasingly important for securities regulators.²⁰ Logically, since the coming into force of Company Law 1993, the Central Committee of the Communist Party of China shifted the corporate objectives of reform in state-owned enterprises (SOEs) from power-delegating and profit-sharing to the establishment of a modern enterprise system, in the context of which company law was a significant factor.²¹ However, in CCL 1993 there was no explicit section allowing for derivative action on the part of shareholders. Article 111, the only provision referring to shareholders' action in CCL 1993, gave shareholders legitimate rights of action but failed to indicate whether derivative action would be permitted. Claimants in a derivative claim can sue based on the corporate right of action, and the right provided in art 111 of CCL 1993 was a typical direct action, rather than being derived from the rights of the company. No substantial or procedural rules have been provided in art 111 for derivative claims to make the claim fundamentally different from shareholders' personal actions, and therefore the article cannot be regarded as a legal basis for derivative action in China.²²

Legislative interpretation completed the positive position of derivative action when Mr Guogang Li, the deputy president of the Supreme People's Court, concluded that courts would accept derivative action in his speech to the Meeting of China's Court Adjudication

14 See *Percival v Wright* [1902] 2 Ch 421 Ch D.

15 *Cooke v Cooke* [1997] 2 BCLC 28 Ch D.

16 Bgb1 I P 2802.

17 See China Securities Regulatory Commission, Statistics Data of September 2012 <www.csrc.gov.cn/pub/zjhpublic/G00306204/zqscyb/201210/t20121017_215899.htm>.

18 See China Securities Regulatory Commission, *China Capital Markets Development Report* (in Chinese) (2008) 269 <www.csrc.gov.cn/pub/newsite/yjzx/cbwzx/ebook/P020090315493496092493.pdf>.

19 *Ibid.*

20 X Huang, 'Shareholders Revolt: The Statutory Derivative Action in China' (Comparative Research in Law and Political Economy Research Paper 49/2009) 5.

21 F X Hong and S H Goo, 'Derivative Action in China: Problems and Prospectus' [2009] *Journal of Business Law* 376, 377.

22 However, there is controversy about this, since it has been argued that the right to derivative action can in fact be implied elsewhere in art 14 (2) of CCL 1993 and art 63 of CCL 1993.

Work for Civil and Commercial Affairs on 11 December 2002. The legitimate position of shareholders in bringing a derivative claim was further confirmed in 'Opinions on Some Issues in Trials for Legal Actions Related to Company Disputes (No 1)' promulgated by the Shanghai People's Court in 2003 when the system was first introduced in China. Further, 'Opinions on Some Issues in Trials for Legal Actions Applied with Company Law', promulgated by the High People's Court of JiangSu Province in 2003, and 'Opinions on Some Issues in Trials for Legal Actions Related to the Company Disputes' promulgated by the Beijing High Court one year later also stipulated the rules for shareholder derivative action. However, these judicial opinions were only for judges to use for reference, and as such they did not have legally binding effects. Due to the disparities in procedural requirements and the lack of experience and common law support, it was still very difficult for shareholders to successfully bring derivative claims.

A few cases need to be discussed here to illuminate the Chinese derivative action system pre-CCL 2006. Despite the fact that derivative action is accepted by both the local courts and wider legislation, not many suits have been brought against companies, especially publicly listed companies, either before or after the formal introduction of the system in CCL 2006. Derivative suits filed by minority shareholders declined due to the fact that there was no legal basis for derivative action. The Shanghai Intermediate People's Court refused to entertain it in *Zhongtian v Bichun* in 1996. The courts first acknowledged a derivative action lawsuit in the case of *Zhangjiagang Fibre Company* in 1994, where the Supreme People's Court was asked for guidance on the question of whether a factory, a Chinese investor in a Chinese-foreign equity joint venture, was able to represent the joint venture in bringing a lawsuit against a vendor to the joint venture. The Supreme Court held that the Chinese factory could exercise the litigation right of the cooperative joint venture. Despite the fact that the case was hedged with many conditions and had limited broader applicability, this was the first specific acceptance of a derivative action where shareholders were given legitimacy to take an action to redress a wrong where a company was unable to do so.²³ The response from the Supreme Court was used and cited in cases such as *Zhejiang Wu Fang Zhai* in 2001, and in decisions by the Guangdong Higher People's Court in *Guangzhou Tianhe Scitech* in 2003. However, because the judgment in the Guangdong *Tianhe Scitech* case was issued after 1 January 2006, CCL 2006 could have been applied although the court chose not to.

Regarding listed companies, three cases have been filed by minority shareholders, without exception due to the misappropriation of companies' capital by controlling shareholders. In the case of *Mr Shao v Zhao Xinxian* (000999), Mr Shao, as a minority shareholder of Sanjiu Medical and Pharmaceutical Co Ltd, filed a lawsuit on 8 April 2003 against Zhao Xinxian, the chair of Sanjiu, due to misappropriations in Sanjiu. The law suit was rejected on the grounds that 'derivative action shall be brought based on the interests of all shareholders and Mr Shao needs the permission of all shareholders for filing the lawsuit', which was impossible in the case of Sanjiu considering the shareholding ownership structure at the time, according to which the two largest shareholders owned more than 70 per cent of the shares and were both involved in the misappropriations. The court denied

23 See *Zuigao Renmin Fayuan Guanyu Zhongnwaibeizi Jingying Qiye Duinai Fasheng Jingji Hetong Jiuwen, Kongzhi He Ying Qiye De Waitang Yu Maifang You Libai Yuansci, Heying Qiye De Zhongfang Yingyi Sheide Mingyi Xiang Renmin Fayuan Qisu Wenti De Fuan* 1994 (Reply from the Supreme People's Court on the issues in whose name shall the Chinese side of the joint venture file a lawsuit to the People's Court when a Chinese foreign equity joint venture has an external controversy over an economic contract, and the foreign side which controls the joint venture has a direct interest relationship with the seller 1994); see also D C Clarke and N C Howson, 'Pathway to Minority Shareholder Protection: Derivative Action in People's Republic of China' in D W Puchniak, H Baum and M Ewing-Chow, *The Derivative Action in Asia: A Comparative and Functional Approach* (CUP 2012) 243, 265.

the enforcing effect of the Corporate Governance Code 2002²⁴ and took a cautious attitude in the absence of litigations on derivative action. In addition to the case of *Sanjin*, derivative action was accepted in another two cases with unknown subsequent developments, including *Henan Lianhua Weijing* (600186)²⁵ and *Shenzhen Xingdu Hotel* (000033).

In order to promote the position of shareholders and improve the incomplete and unsatisfactory legislative approach in CCL, the CCL 2006 provides shareholder with cumulative voting rights,²⁶ the right to request the companies to repurchase their shares at a reasonable price²⁷ and the right of viewing important company documents.²⁸ Most innovatively, the provision for statutory derivative action was introduced in art 152 of CCL 2006 as one of the key amendments introduced to enhance shareholder protection:

where a director or senior manager is under the circumstance as stated in Article 150 of the law (where a director, supervisor or senior manager violates the law, administrative regulations or the article of association during the course of performing his duty), the shareholders of the limited liability company or joint stock limited company separately or aggregately holding one per cent or more of the total shares of the company for at least 180 consecutive days before the commencement of the action may require the board of supervisors or the supervisor of a limited liability company with no board of supervisor in writing to file a lawsuit in the People's Court.

Besides this, the provision of a mechanism for shareholder derivative action can also be found in Chinese Securities Law 2005, where:

shareholder derivative actions can be brought by shareholders when directors, supervisors, senior managerial staff or shareholders holding more than five per cent or more of the company shares have breached the restriction on trading in company shares during a six-month period.²⁹

There was an increase in the number of lawsuits brought by minority shareholders based on s 152 after the enforcement of the derivative claim action in CCL 2006.³⁰ However, most of these lawsuits have been focused on limited liability companies, with a striking absence of actions targeted at companies limited by shares.³¹ As for listed companies, there has only been one lawsuit brought by shareholders and subsequently accepted by the Shandong Higher People's Court on 11 December 2009. The *Sanlian Shangshe* case (2009), as the first shareholder derivative lawsuit towards a listed company after the enforcement of CCL 2006, has important practical significance. However, the case is still *sub judice* and the filing of the case is relatively slow. Unlike classic derivative action cases, this case involves an attempt by the plaintiffs, 78 shareholders of Shanlian Shangshe, to bring a derivative

24 See Code of Corporate Governance for Listed Companies 2002, art 4.

25 For the case report see www.sse.com.cn/cs/zhs/scfw/gg/ssgs/2004-07-23/600186_20040723_1.pdf.

26 CCL 2006, art 106.

27 CCL 2006, art 75.

28 CCL 2006, art 34.

29 Chinese Securities Law 2005, art 47.

30 Y Ji and L Zhu, 'Dissuasions of Jurisdictional Issues of Shareholder Derivative Lawsuit in China (*Qiantan Gudong Paisheng Susong Anjian de Guanxiaguan Wenti*)' (2008) available via China Court website <<http://old.chinacourt.org/html/article/200812/09/334742.shtml>>; see cases such as *Li Xiaozhang and Twenty other shareholders v Xiao Wuyong & Zhang Dingzhong re: Nanchuan Municipal Hardware Infrastructure Electric Chemical Industry Company Limited* (2006); *Dong Fengchang v Fang Yishu re: Shanghai Zhongjian Enterprise Company Limited* (2008); *Zhu Yongjun and Twenty other Shareholder Plaintiffs v Liu Huanren, Zhu Yongjun and Ma Zhonghua re: Shizuishan Municipal Autonomous Region Shizuishan Municipal Intermediate People's Court* (2009).

31 See D C Clarke and N C Howson, 'Pathway to Minority Shareholder Protection: Derivative Action in People's Republic of China' in Puchniak et al (n 23) 275-76.

claim against the former controlling shareholder. From discussions on the reality of applications of derivative action pre- and post-CCL 2006, it is obvious that derivative action in China has been used in a very limited fashion by minority shareholders, despite extensive reports of the power abuse by controlling shareholders and directors. In the next sections, a few factors that hinder the application of the system and related proposed reform suggestions will be discussed with reference to two corporate governance theories.

3 Factors hindering the effective enforcement of derivative action in China

3.1 SHAREHOLDING PERCENTAGE AND SHAREHOLDING PERIOD

The *locus standi* requirements were imposed for shareholders who want to put forward a derivative claim in China in order to prevent unmeritorious suits.³² Based on art 2 of CCL 2006, CCL only deals with two types of company including limited liability companies (*youxian zeren gongsi*) and companies limited by shares (*gufen youxian gongsi*). A limited liability company is normally intended for a much smaller and more closely connected group of investors and they are recognised as 'private' or 'closed' companies. On the other hand, companies limited by shares are the equivalent of Delaware corporations or the German *Aktiengesellschaft* which may be traded on the stock exchange. Under art 152 of CCL 2006, the issues are treated differently depending on the type of company involved. In the case of limited liability companies any shareholder has standing to sue, while standing requirements are imposed for joint stock limited liability companies. There are advantages in adopting this approach to the issue of standing. It is argued that the 'plight of minority shareholders in the limited liability company is generally graver than that of their counterparts in the joint stock limited liability company'.³³ Shareholders in limited liability companies are in a comparatively disadvantaged position since they lack a liquid market, while shareholders in joint stock limited liability companies can choose to leave the company by selling their shares on the stock exchange market. Furthermore, the regulations issued by authorities such as the China Securities Regulatory Commission and the Shenzhen and Shanghai stock exchanges have offered joint stock limited liability companies adequate monitoring, and, as a result of this monitoring, a better corporate governance framework within these companies.³⁴ Also, the abuse of derivative action in limited liability companies is supposedly rare.³⁵ Therefore, the need to establish a derivative claim system to support shareholders and improve corporate governance seems greater in limited liability companies.³⁶

There are deficiencies in the requirements for standing to sue. The 1 per cent shareholding requirement was imposed on shareholders in joint stock companies who want to bring a derivative claim, but not in limited liability companies. Based on a survey conducted by Yang in 2006 examining the shareholding ownership structure of listed companies in China with a highly concentrated shareholding ownership, the proportion of

32 The requirements, however, are not just limited to China; for discussions on requirements in the UK, see V Joffe, D Drake, G Richardson, D Lightman and T Gollingswood, *Minority Shareholders: Law Practice and Procedure* (4th edn, OUP 2011) ch 3.

33 H Huang, 'The Statutory Derivative Action in China: Critical Analysis and Recommendations for Reform' (2007) 4(2) *Berkeley Business Law Journal* 227, 237.

34 Evidence of a better corporate governance framework can be found in art 123 of CCL 2006, requiring a listed company to have independent director(s).

35 R B Thompson and R S Thompson, 'The Public and Private Faces of Derivate Lawsuits' (2004) 57 *Vanderbilt Law Review* 1747, 1784–85; see also X N Li, *A Comparative Study of Shareholders' Derivative Action: England, the United States, Germany and China* (Dordrecht: Kluwer 2007).

36 J Deng, 'Building an Investor-Friendly Shareholder Derivative Lawsuit System in China' (2005) 46 *Harvard International Law Journal* 347; for the relevant empirical research see Ramsay and Saunders (n 3).

shares held by the largest shareholder is in excess of 45 per cent, the second largest is about 5 per cent, while the third is about 3 per cent. No individual shareholder held more than 1 per cent of any company's shares.³⁷ Therefore, the chance of owning more than 1 per cent of the shares is very small for individual shareholders, apart from the top three shareholders. Besides, it has been shown that the average shareholding of one institutional shareholder of a company was approximately 1.4 per cent,³⁸ and the data indicate that it is unlikely that individual minority shareholders will meet the entry level requirement. Furthermore, the minimum capital requirement for being a listed company is RMB 30 million according to Chinese security law³⁹ and, therefore, shareholder(s) who want to bring a derivative law suit must individually or collectively own at least RMB 0.3 million of shares to be qualified plaintiffs of a derivative action. However, the 2011 annual report of the Shanghai Stock Exchange reveals that this is true for only 4.69 per cent of the individual investors in the Shanghai Stock Exchange.⁴⁰ It is also very hard to call for a collective action for shareholders in listed companies with dispersed shareholding ownership, considering the fact that this can be a very time-consuming and expensive process with diversified opinions from different shareholders, and even a lack of knowledge about the operation of the corporation.

The percentage requirement means that the legitimate number of shareholders varies dramatically between large listed companies and comparatively smaller companies. Consequently, collective action from shareholders in bringing a claim against misbehaving directors will be much harder in large corporations. Because of this, Zhu and Chen proposed a replacement of the percentage requirement with a requirement based on the value of shares.⁴¹ Alternatively, others have suggested that a lower percentage of the shareholding in listed companies in China should be necessary in order to qualify for a derivative claim.⁴² In the absence of systemic restrictive procedures for derivative action in Chinese civil procedure law, a quantitative requirement should be imposed before a derivative claim to avoid an excessive amount of litigation or malicious or frivolous law suits from shareholders. Therefore, in the author's view, a requirement based on the value of the share recruitments seems more enforceable, encouraging shareholders who have standing to make a sincere claim.

Besides this, the shareholders of joint stock limited companies are required to hold their shares for at least 180 consecutive days before the action starts.⁴³ This standing rule is designed to stop people from buying shares with the intention of bringing malicious litigation, and to prevent excessive workload for directors, supervisors and senior managers in dealing with claims. With a very immature stock market, the only aim of most Chinese shareholders is to get immediate returns from their investments through share transactions. The understanding of shareholders of the operations and management policies involved in running corporations is either very limited or non-existent. Statistics showed that the

37 J Yang, 'The Role of Shareholders in Enforcing Directors' Duties: A Comparative Study of the United Kingdom and China (Part 2)' (2006) 7 *International Company and Commercial Law Review* 381, 390.

38 Y Zeng and J Zhang, 'The Dark Side of Institutional Shareholder Activism in Emerging Market Evidence from China's Non-Tradable' (2011) 40 *Asia-Pacific Journal of Financial Studies* 240

39 China Security Law 2005, art 16.

40 Shanghai Stock Exchange, *Shanghai Stock Exchange 2011 Annual Report* (Shanghai Stock Exchange 2011) 346.

41 H C Zhu and G Q Chen, 'China Introduces Statutory Derivative Action' (2005) 24 *International Financial Law Review* 21, 22.

42 J H Liu, *On the Shareholder's Derivative Action* (Beijing: Chinese People's Public Security University Publisher 2008), ch 3.

43 Article 152 CCL 2006; see also the Supreme People's Court Regulations on the Application of Company Law of the People's Republic of China 2006.

average shareholding period on the securities market in 2005 was less than 120 days.⁴⁴ Therefore, the 180-day requirement does not appear to be practical in the unique Chinese context. Logically, it could be suggested that the shareholding period might be shortened to make the derivative claim an effective way of obtaining shareholder remedies.

The calculation of the 180-day period is also problematic. There are generally three interpretations of the term '180 consecutive days': (1) from the date of purchasing the shares to the date of the misconduct of the directors, supervisors or senior management; (2) from the date of purchasing the shares to the date of bringing the derivative action where the misconduct of the directors, supervisors or senior management happens *after* the date of purchasing the shares; or (3) from the date of purchasing the shares to the date of bringing the derivative action where the misconduct of the directors, supervisors or senior management happens *before* the date of purchasing the shares. In a positive light, this was interpreted by the Supreme Court in art 4 of the first section of the Supreme People's Court Regulations on the Application of Company Law of the People's Republic of China 2006. It is explained that the term '180 consecutive days' in art 152 of CCL 2006 refers to the period of time from the date of purchasing the shares to any date before or on which the shareholders bring the derivative action. Despite the fact that this interpretation makes the application of art 152 easier in practice, the cogency or legal force of this judicial interpretation, compared with the law adopted at the National People's Congress, is much weaker and less authoritative.

Achieving eligibility right is fundamental to making derivative action work effectively in China. In comparison with some jurisdictions, it is arguable that the rule on eligibility in China seems to be both under-inclusive and over-inclusive, since it may restrict certain meritorious claims while at the same time failing to discourage vexatious actions.⁴⁵ Compared with the law in Australia and Canada,⁴⁶ the eligibility requirements in China seem overly restrictive. According to the Australian Corporations Act 2001, the scope of the term eligible defendant is extremely broad and includes 'a member, former member, or person entitled to be registered as a member, of the company or of a related body corporate; or an officer or former officer of the company'.⁴⁷ Past members were included because they might have been compelled to terminate their relationship with the company due to the underlying dispute, while members were included since they may have been adversely affected by the directors' misconduct.⁴⁸ On the other hand, stipulations in CCL 2006 regarding eligible claimants appeared to be over-inclusive, making it hard to discourage unmeritorious claims. This is unlike the requirement of 'contemporaneous ownership of shares' in the US, under which the claimants 'must have been a shareholder or member of the corporation or association at the time of the transaction of which complaint is made or that the share or membership, if acquired thereafter, devoid by operation of law'.⁴⁹ Therefore, potential shareholders who are qualified to bring a claim need to be shareholders

44 X Jin, *Shareholding Structure and Corporate Governance in Listed Companies in China* (Beijing: China Finance Press 2005), 143–44.

45 Huang (n 33) 242.

46 See Canada Business Corporations Act, RSC 1985, ss 238–39.

47 Australian Corporations Act 2001, s 236(1)a(i) and (ii).

48 Corporate Law Economic Reform Program Bill, HR 71, 105th Congress. Explanatory Memorandum, paras 6.26–28 (1998) <www.comlaw.gov.au/Details/C2004B00269/Explanatory%20Memorandum/Text>.

49 Federal Rules of Civil Procedure, r 23.1.

at the commencement of the suit and during the pendency of the suit.⁵⁰ This requirement in US law is designed to ‘prevent what has been considered an evil, namely, the purchasing of shares in order to maintain a derivative action designed to attach a transaction which occurred prior to the purchase of the stock’.⁵¹ Conversely, under CCL 2006, unjustifiable shareholders are able to bring claims by acquiring shares after finding a potential opportunity for bringing derivative action. Neither under-inclusive nor over-inclusive requirements for eligibility of defendants will create a healthy and workable environment for a better enforcement of CCL 2006, at least as far as the provisions on derivative actions are concerned.

3.2 SCOPE OF DERIVATIVE ACTION STANDING

The range of reasons that shareholders can use to bring derivative actions seems very broad based on stipulations in CCL 2006, which may include any violations of ‘laws, administrative regulations or the articles of association during the course of performing his duties’. Therefore, the shareholders will be able to challenge any breach of the duties by directors according to law, regulations or articles of association. These may include breaches of fiduciary duties or duty of diligence with broad scope in the new CCL.⁵² Article 5 of CCL 2006 states that ‘a company must, when engaging in business activities, abide by the laws and administrative regulations, observe social morals and business ethics, be in integrity and good faith, accept regulation of the government and the public, and undertake social responsibilities’. Comparatively modern terms such as ‘business ethics’ and ‘social responsibilities’ are introduced in the general provisions of CCL for the first time. It is implied in art 5 that apart from the interests of shareholders, employees and other stakeholders, the performance and activities of the company have a deep impact on the market’s economic rules and public social interests. Therefore, when company directors and supervisors pursue company interests for their shareholders, they have to be socially responsible and responsible to internal and external stakeholders.⁵³ As far as legislative tenets are concerned, it is stipulated in art 1 of CCL 2006 that this legislation is enacted in order to ‘standardise the organisation and activities of companies, to protect the legitimate rights and interests of companies, shareholders and creditors, to maintain socio-economic order and to promote the development of the socialist market economy’.⁵⁴ However, these socially related responsibilities and legislative tendencies in protecting creditors’ interests were first introduced as a guidance principle for future detailed stipulations regarding directors’ duties, such as their detailed responsibilities towards various stakeholders, their duties towards stakeholders in realising these responsibilities, and the liability of corporations and directors if they are in breach of these duties. Directors’ duties relating to the social aspects required

50 See *West v West* 126 FRD 82 (ND Ga 1992); Compare *Brown v Ferro Corp* 763 F2d 798 (6th Cir) cert denied 474 US 947 (11985); *Rotenberg v Utd Brands Co* 574 F2d 1295 (2d Cir 1997); *deHaas v Empire Petroleum Co* 300 F Supp 834 (1969) & 435 F2d 1223 (10 Cir 1997); *Drachman v Harvey* 435 F2d 722 (2d Cir 1971) modified *en banc* 453 F2d 736 (1972).

51 *Rosenthal v Barry Biscuit Corp* 60 A 2d 106, 111 (Del Ch 1948) see also *Anadarko Petroleum Corp v Panhandle Eastern Corp* 545 A.2d 1171 (Del 1988); *Haff v Kerkorian* 324 A 2d 215 (Del Ch 1974) *affd in part, revd in part* 347 A 2d 133 (Del 1975); see also C Ferrara, K T Abikoff and L L Gansler, *Shareholder Derivative Litigation: Besieging the Board* (New York: Law Journal Press 2013) ss 4.01–02.

52 CCL 2006, arts 148 and 149.

53 J Zhao, ‘The Regulation and Steering of Corporate Social Responsibility in China: Stories after the Enforcement of Chinese Company Law’ (2011) 32 *International Commercial and Company Law Review* 400; see also J Zhao, ‘The Harmonious Society, Corporate Social Responsibility and Legal Responses to Ethical Norms in Chinese Company Law’ (2012) 12 *Journal of Corporate Law Studies* 163; J Zhao, *Corporate Social Responsibility in Contemporary China* (Cheltenham: Edward Elgar Publishing, forthcoming).

54 CCL 2006, art 1.

in CCL 2006 are immature, unclear and extremely difficult to enforce.⁵⁵ It is overly demanding for directors to enforce these duties in accordance with the legislative principles, and the danger of being held liable for negligence of these responsibilities could be an onerous burden for directors and deter them from taking risky decisions.

Furthermore, the duty of diligence and duty of loyalty have been introduced into CCL. Directors might be held responsible for breaching these duties simply by making bad business decisions and these onerous burdens will have a negative impact on directors when they are trying to make risky management decisions for the long-term interests of corporations. If shareholders are allowed to bring a claim against directors based on infringements of such unclear duties, this may lead to a large number of malicious suits. To counter this defect, efforts should be made to make these provisions regarding directors' duties more accessible and easy to understand for directors, especially those in small and medium-sized businesses. The improvement can be achieved by various measures, including clearer interpretations, broader enrichment with systematic codification of relevant legislation on directors' duties, and more judicial regulations and explanations from the People's Supreme Court. Also, in terms of the enforcement of directors' duties for more socially responsible corporations, establishing a sound corporate governance model with a clearer corporate objective is critical for the efficient and sustainable performance of Chinese companies.

It is also recommended that Chinese courts should adopt the concept of the business judgment rule, which was designed in the US to 'stimulate risk taking, innovation and other creative entrepreneurial activities'⁵⁶ with the purpose of striking a good 'balance between the need to ensure that directors comply with the duty of diligence and the need to allow directors to take certain risks in the exercise of their discretion'.⁵⁷ The business judgment rule was introduced more than 180 years ago as a common law standard.⁵⁸ It has been regarded as the principle that directors can employ to shield their decisions from judicial scrutiny.⁵⁹ The most distinctive contribution of the business judgment rule is that it prevents the judiciary from meddling in managerial decisions.⁶⁰ In this sense, the rule may be helpful in protecting directors from personal liability for claims made against them because of errors of judgment or wrong business decisions, protecting them against shareholders' derivative

55 L W Lin, 'Corporate Social Responsibility in China: Window Dressing or Structural Change' (2010) 28 *Berkeley Journal of International Law* 64.

56 See s 4.01 of American Law Institute, *Principles of Corporate Governance: An Analysis and Recommendations* (Proposed Final Draft, ALI 1992); see also J E Kerr, 'Sustainability Meets Profitability: The Convenient Truth of How the Business Judgment Rule Protects a Board's Decision to Engage in Social Entrepreneurship' (2007) 29 *Cardozo Law Review* 623; C X Weng, 'Assessing the Applicability of the Business Judgment Rule and the "Defensive" Business Judgment Rule in the Chinese Judiciary: A Perspective on Takeover Dispute Adjudication' (2010) 34 *Fordham International Law Journal* 124; R Sprague and A J Lyttle, 'Shareholder Primacy and the Business Judgment Rule: Arguments for Expanded Corporate Democracy' (2010) 16 *Stanford Journal of Law, Business and Finance* 1; see also *Percy v Millaudon* (1829) 8 Mart (ns) 68; *Revlon Inc v MacAndrews & Forbes Holdings Inc* (1986) 506 A2d 173 and more recently cases like *Gantler v Stephens* (Del 2009) 965 A2d 695.

57 F Ma, 'The Deficiency of Derivative Action in China' (2010) 31 *Company Lawyer* 150, 153; see also D M Branson, 'A Business Judgment Rule for Incorporating Jurisdictions in Asia' (2011) 23 *Singapore Academy of Law Journal* 687; J H Farrar, 'Directors' Duties of Care: Issues of Classification, Solvency and Business Judgment and Dangers of Legal Transplant' (2011) 23 *Singapore Academy of Law Journal* 745.

58 See S S Arusht, 'The Business Judgment Rule Revisited' (1979) 8 *Hofstra Law Review* 93, 93.

59 See J Hinsey, 'Business Judgment and the American Law Institute's Corporate Governance Project: The Rule, the Doctrine and the Reality' (1985) 52 *George Washington Law Review* 609, 610.

60 C X Weng, 'Assessing the Applicability of the Business Judgment Rule and the Defensive Business Judgment Rule in the Chinese' (2010) 34 *Fordham International Law Journal* 124.

actions based on their negligence.⁶¹ Both academics and practitioners in China have realised the importance of the concept of business judgment, and it is suggested that Chinese judges should reference the principle in managing their discretions when they judge whether directors are in breach of the duties imposed in CCL 2006.⁶²

3.3 FUNDING DERIVATIVE ACTIONS IN CHINA AND INCENTIVES FOR CHINESE SHAREHOLDERS TO BRING ACTIONS

The cost of litigation has always been a key issue in derivative action, since the related rules will have a direct impact on the popularity of the suit. Private shareholder actions have been regarded as a major feature of corporate governance in the US, but this is not the case in the UK because of the fact that the rule on litigation costs promotes the launching of law suits in the US but discourages the same in the UK.⁶³ The different treatment of litigation fees can group the jurisdictions adopting derivative claims into three streams. In the US, it is the litigant's responsibility to fund their litigation expenses without considering the results of the litigation, while the attorney's fees are contingent only on successfully litigated or settled cases.⁶⁴ The traditional regime adopted in the UK is the 'loser pays' regime, where fees are typically assessed against the loser on 'a standard basis' which covers litigation costs 'proportionately and reasonably incurred and proportionate and reasonable in amount'.⁶⁵ However, conditional fee arrangements were introduced in England as a result of the reform of civil litigation funding.⁶⁶ In this context, shareholders who want to bring derivative claims could adopt 'no win no fee' arrangements with their lawyers. Thirdly, in Japan, the derivative claim cost is fixed at 8200 Japanese yen, by deeming derivative actions to be non-property claims.⁶⁷ This followed a new interpretation given by the Tokyo High Court on 30 May 1993.⁶⁸

It is argued that current Chinese law and regulations on both civil litigation filing fees and attorneys' fees effectively deter shareholders from bringing derivative claims on behalf of corporations.⁶⁹ As far as filing fees are concerned, the rule on these fees is regulated by the Measure on Payment of Litigation (*Susong Fei Jiaona Banfa*) 2007. According to art 29 of

61 M B Hemraj, 'The Business Judgment Rule in Corporate Law' (2004) 5 *International Company and Commercial Law Review* 192, 198.

62 F Deng, 'A Legal Analysis of Leadership Accountability: An Approach from the Director's Duty of Care in Corporate Law' (2007) 28 *Social Science in China* 3.

63 J Armour, B Black, B Cheffins and R Nolan, *Private Enforcement of Corporate Law: An Eremitical Comparison of the US and UK* (Preliminary Draft of Law and Economic Working Paper No 89, University of Texas School of Law 2007).

64 *Alyeska Pipeline Service Co v Wilderness Society* 421 US 240 (1975); *Hall v Cole* 412 US 1 (1974); *Mills v Electric Auto-Lite Co* 396 US 375 (1970); *Duban v Diversified Mortgage Investors* 87 FRD 33 (SD NY 1980); see also D A DeMott, *Shareholder Derivative Action: Law and Practice* (Thompson 2003) s 6.12.

65 CPR 1998, SI 1998/3132, 44.5; *Lownds v Home Office* [2002] 1 WLR 2450, 2453.

66 J Peynser, 'A Revolution by Degrees: From Costs to Financing and the End of the Indemnity Principle' [2001] *Web Journal of Current Legal Issues*; see also A Hogan, 'Conditional Fees: Problems Solved and Problems Yet to Come' (2006) 1 *Journal of Personal Injury Law* 40.

67 Commercial Code of Japan 1993, art 267(4).

68 See *Asai v Inasaki* (or *Nikko Securities case*) Tokyo District Court, 101 Shiryoban Shoji Homu 37, revd, Judgment of 30 March 1993, 109 Shiryoban Shoji Homu 70. It is worth mentioning that this interpretation of the old code provision was confirmed by the Supreme Court on 10 October 2002; see T Fujita, 'Transformation of the Management Liability Regime in Japan in the Wake of the 1993 Revision' in H Kanda, K S Kim and C J Milhaupt (eds), *Transforming Corporate Governance in East Asia* (Routledge 2008) 15, 17.

69 F P Meng, 'Funding Derivative Action in China: Lessons from *Wallersteiner v Moir* (No 2) for the Court' (2010) 31 *Company Lawyer* 29, 29; see also Z Zhang, 'Making Shareholder Derivative Action Happen in China: How Should Lawsuits be Funded' (2010) 40 *Hong Kong Law Journal* 523.

the Measure, a major part of the fee the litigation bears is the case acceptance fee, which is based on the loser pays rule. Similar to the litigation calculation rules adopted in Japan before the 1993 Commercial Code, a sliding scale system is adopted. This means that the litigation fee is charged in proportion to the amount of damage the defendants seek to claim. According to art 13 (1) of the Measure, RMB 50 is charged for damages of no more than RMB 10,000. After this fixed amount, different percentages of the damages sought will be charged for different damage amounts. However, the sliding-scale system has been historically proven to be a barrier to shareholder derivative claims in Japan.⁷⁰ The actual effect of the pre-1993 Japanese litigation fee principles was to discourage all suits due to the high litigation fees, which are not recoverable unless the shareholder eventually wins the case.⁷¹ Statistics show that there were less than 20 derivative action suits between 1950 and 1990, but there was a dramatic increase after 1993 to hundreds of law suits each year.⁷² Therefore, in the author's opinion, a reform should also be adopted in China to fix the charge of the litigation cost, since the amount of damages to be paid to the suing shareholder is normally difficult to determine. Therefore, a relatively lower fee of RMB 1000 Yuan (equivalent to 8200 yen) should be charged.

As for the lawyers' fees, two options are available according to the Measures for the Administration of Attorneys' Fees 2006: the 'contingency fee arrangement', where the fees are fixed at a percentage no higher than 30 per cent of the amount of the damages claimed as the result of a successful litigation;⁷³ and the 'reasonable fee arrangement', where the fees are charged on the basis of the subject-matter of the case or of the time taken.⁷⁴ However, neither arrangement seems fair or reasonable for shareholders, or gives them sufficient incentive to bring a derivative action. If the litigation is successful, the benefits awarded accrue to the company. The defendant shareholder would have to personally fund the lawyer's fee because of the service contract between the lawyer and the shareholder. Under the contingency fee arrangement, the shareholder would be personally liable for up to 30 per cent of the damages recovered if the litigation is successful. Paradoxically, if the litigation is not successful, the shareholder might be in a better position because of being free from responsibility for the lawyer's fee.

A more sensible fee arrangement⁷⁵ is important for a more enforceable derivative claim system and better corporate governance in China. The author agrees with the 'indemnity order' enforced in the UK,⁷⁶ which was first made legitimate in *Wallersteiner v Moir (No 2)*⁷⁷ where Lord Denning asserted that a company should indemnify a shareholder defendant in a derivative suit since the shareholder acts on behalf of the company and the company is the direct beneficiary. The shareholders who bring the claim should be treated as trustees or agents who act on behalf of the beneficiary or the principal. Lord Denning suggested a more reasonable procedural rule when Mr Moir's plight found sympathy with the Court of Appeal:

70 M Blomstrom and S La Croix, *Institutional Change in Japan* (Routledge 2006) 241; see also M West, 'The Pricing of Shareholder Derivative Action in Japan and United States' (1994) 88 *Northwestern University Law Review* 1436.

71 C J Milhaupt and M D West, *Economic Organizations and Corporate Governance in Japan: The Impact of Formal and Informal Rule* (OUP 2004) 21.

72 M D West, 'Why Shareholders Sue: The Evidence from Japan' (2001) 30 *Journal of Legal Studies* 351, 352.

73 Measures for the Administration of Attorneys' Fee 2006, arts 11–13.

74 *Ibid* art 10.

75 Meng (n 69) 29–30.

76 *Ibid* 29.

77 *Wallersteiner v Moir (No 2)* [1975] QB 373; see also *Masri v Consolidated Contractors International Co SAL* [2011] EWHC 1024; *Air Canada v Secretary of State for Trade (No 2)* [1983] WLR 494.

The minority shareholder should apply ex parte to the master for directions, supported by an opinion of counsel as to whether there is a reasonable case or not. The master may then, if he thinks fit, straightaway approve the continuance of the proceeding until close of pleadings, or until after discovery or until trial . . . The master need not, however, decide it ex parte . . . The master should simply ask himself: is there a reasonable case for the minority shareholder to bring at the expense (eventually) of the company? If there is, let's go ahead.⁷⁸

It is normally non-arguable that, even if the derivative action is successful, the recoveries accrue to the company as a whole with the shareholder who brought the claim receiving only a small percentage of the benefit. The fact that other shareholders may have a 'free ride' and directly benefit from derivative claims brought by the defendant shareholder, who did all the hard work, will make defendant shareholders less willing and give them less incentive to bring such actions in the future.

4 Corporate governance theory and the establishment of a more efficient derivative action scheme in China

Effective from 2006, the derivative claim action system embedded in CCL was designed to improve the Chinese corporate governance system and provide a weapon against insider and controlling shareholder abuse in newly corporatised listed companies in China.⁷⁹ Since the enforcement of CCL 2006, the demand for minority shareholder protection has increased as the result of demand from several directions. With the increasing number of dividable shareholders, largely due to the expansion of the securities market, the rights and liabilities of investors are promoted by the public media in the aftermath of corporate scandals and market turbulence.⁸⁰ The focus on the development of the stock market in order to ease the pressures on loan capital from banking finance is making shareholders' rights and remedies increasingly important.

The conversion of SOEs enables the government to be less concerned about providing guidance and privileges to SOEs, and by extension the state as the biggest shareholder in SOEs.⁸¹ China has experienced an unprecedented wave of corporatisation and privatisation since implementing CCL 1993.⁸² The corporatisation of business organisations with decreased government interference under CCL 2006, in comparison with CCL 1993, means that corporate governance theory is closely related to the protection of the shareholder in China.

Systematic corporate governance reform is becoming increasingly important in the emerging Chinese market, which is moving towards a more free-market system.⁸³ Shareholder derivative claims are always used as a 'case study' for the relationship between

78 *Wallersteiner v Moir (No 2)* [1975] QB 373 AT 392.

79 D C Clark and N C Howson, 'Pathway to Minority Shareholder Protection: Derivative Action in the People's Republic of China' in D Puchniak, H Baum and M Ewing-Chow (eds), *The Derivative Action in Asia: A Comparative and Functional Approach* (CUP 2012) ch 6, 1.

80 X Tang, 'Protecting Minority Shareholders in China: A Task for Both Legislations and Enforcement', ch 9 in Kanda et al (n 69).

81 *Ibid.*

82 Y Wei, 'The Development of Securities Market and Regulations in China' (2005) 27 *Loyola of Los Angeles International and Comparative Law Review* 479.

83 S C Jain, *Emerging Economies and the Transformation of International Business* (Cheltenham: Edward Elgar Publishing 2006) 383–86.

institutions and the norms of corporate governance.⁸⁴ Chinese government officials and commentators believe that shareholder derivative claim actions will promote better corporate governance in China,⁸⁵ and it is an ‘almost unanimous understanding’ that China has to institute a system for shareholder derivative suits.⁸⁶ Conversely, however, corporate governance theories could be adopted as an effective way of illuminating an improved institutionalised derivative claim system. The lack of a shareholder remedy contributed to the inefficient corporate governance system in China, especially for minority shareholders.⁸⁷ Proponents of derivative action describe it as ‘the chief regulator of corporate management’.⁸⁸ It plays a significant role in deterring managerial misconduct, and thereby reduces the agency cost inherent in corporate management.⁸⁹ It is predicted that the adoption of the actions will be ‘a major development of Chinese company legislations’ with ‘far reaching implication for corporate governance’.⁹⁰ However, the effectiveness of transplanting the derivative action rule into China may not always lead to the intended consequences and results.

Corporate governance theories, including agency theory and path dependence theory, can help to explain the Chinese derivation action with unique ‘Chinese characteristics’. China’s economy is in a critical transitional phase with changes in the enterprise system and the securities market. This places the Chinese corporate governance system in a unique position to adopt the derivative action rule. The company law system in China, including civil procedural law, economic transformation, current legal enforcement problems, qualification of practitioners and judges, and Chinese traditions and history, will have to be considered collectively with the corporate governance system to determine the effectiveness of adopting a derivative action system.

‘The judiciary is weak, unsophisticated and riddled with corruption’ in China.⁹¹ Furthermore, Chinese judges are often criticised for their inadequate qualifications, poor training and lack of knowledge and skills. The reputation and authority of the judiciary has been negatively compromised as a result of corruption and political interference in judicial processes. *Guanxi* and *renqing* make the supervisory board, which does not have power over the managerial board and has no remit to elect or dismiss managerial directors, even less efficient (*guanxi* – close personal relationships or connections, which are established connections to secure benefits – and *renqing* – getting benefits or advantages from a relationship – are major elements in Chinese commercial life). Institutional shareholders

84 See examples such as: D C Clark, *The Ecology of Corporate Governance in China* (George Washington University Law and Legal Theory Working Paper No 433 2008); A Keay, ‘The Ultimate Objective of the Public Company and the Enforcement of the Entity Maximisation and Sustainability Model’ (2010) 10 *Journal of Corporate Law Studies* 369; S P Ferris, T Jandik, R M Lawless and A Makhija, ‘Derivative Lawsuits as a Corporate Governance Mechanism: Empirical Evidence on Board Changes Surrounding Filings’ (2009) 42 *Journal of Financial and Quantitative Analysis* 143; A S Ginevri, ‘The Rise of Long-Term Minority Shareholders’ Rights in Publicly Held Corporations and Its Effect on Corporate Governance’ (2011) 12 *European Business Organization Law Review* 587.

85 Deng (n 36) 349.

86 G Yu, ‘Using Western Law to Improve China’s State Owned Enterprise: Of Takeover and Securities Fraud’ (2004) 39(2) *Valparaiso University Law Review* 339, 340.

87 G Yu, ‘Towards an Institutional Competition Model of Comparative Corporate Governance Studies’(2003) 6 *Journal of Chinese and Comparative Law* 31, 56.

88 *Cohen v Beneficial industrial Corp* 337 US 541 (1949) 548.

89 J Coffee and D Schwartz, ‘The Survival of the Derivative Suit: An Evaluation and a Proposal for Legislative Reform’ (1981) 81 *Columbia Law Review* 261.

90 Huang (n 33) 242.

91 Z Zhang, ‘The Shareholder Derivative Action and Good Corporate Governance in China: Why the Excitement is Actually for Nothing’ (2011) 28(2) *UCLA Pacific Basin Law Journal* 174, 175.

and other non-controlling shareholders in China, who may be able to bring such actions, will normally be reluctant to do so because of Guanxi and *renqing* due to the controlling shareholders and management. These factors make the introduction of path dependence theory necessary as a theoretical base for the unique problems of establishing and enforcing an efficient derivative action system in China.

The Chinese corporate governance system is a control-based model, in which the controlling shareholders 'tightly control the list companies through concentrated ownership [and] management friendly boards'.⁹² If derivative action is to play any role, the question is what, if any, are the net benefits of these actions for companies, shareholders and stakeholders from a social or economic perspective. It is important to introduce agency theory to the discussions of Chinese shareholding structure and the problems of derivative action in order to make this legal scheme more enforceable and popular, especially for public companies, and reduce the agency cost. With the unique characteristics of Chinese SOEs and the close *guanxi* between government and corporations, derivative claims are a reasonable remedy and provide an alternative way of protecting the interests of shareholders, particularly minority shareholders.

There is a strong need for China to adopt a workable derivative action system due to the current context of rapid economic transformation, as well as a high incidence of directors and majority members harming the interests of the company or the minority shareholders. In the following section, corporate governance theories will be used to discuss the possible ways of enabling legislators and regulators to make derivative action rules accessible, enforceable and efficient for Chinese investors within the unique Chinese system.

4.1 AGENCY THEORY

Derivative claims by shareholders are closely related to the question of control of the corporate form, and of the steps being taken on behalf of shareholders to redress the imbalance in modern corporations between the control exercised by directors and that exercised by shareholders.⁹³ It is argued that derivative actions generate agency costs, and the risk and fear of being sued might hinder corporate directors from accepting the position in the first place.⁹⁴ On the other hand, derivative actions can reduce agency cost as the claims work to 'deter mismanagement by imposing the threat of liability and therefore align the interests of managers and shareholders'.⁹⁵ The adoption of the derivative claim principle will increase the costs that companies have to incur in order to attract good directors. A manager's net return from his job must equal some 'reservation' level for him to be willing to work for the company based on market forces.⁹⁶

However, it is necessary to expand the scope of shareholder remedies towards a derivative action system because of the difficulty of winning a proxy fight by getting sufficient numbers of shareholders at the requisite meeting, regardless of the fact that

92 Q Liu, 'Corporate Governance in China: Current Practices, Economic Effects and Institutional Determinants' (2006) 52 *CESifo Economic Studies* 415, 429.

93 D McDonough, 'Proposed New Statutory Derivative Action: Does it Go Far Enough?' (1996) 8 *Bond Law Review* 47.

94 R R Kraakman, P Davies, H Hansmann, G Hertig, K J Hopt, H Kanda and E B Rock (eds), *The Anatomy of Corporate Law: A Comparative and Functional Approach* (OUP 2004) s 2.2.1.

95 I Ramsay 'Corporate Governance, Shareholder Litigation and the Prospects for a Statutory Derivative Action' (1992) 15 *University of New South Wales Law Journal* 149, 152.

96 A Reisberg, *Derivative Actions and Corporate Governance Theory and Operation* (OUP 2007) 48.

shareholders have their own contractual remedies which manifest themselves in positive control.⁹⁷ Under these circumstances, shareholders must revert to negative legal control measures to redress the imbalance in control between the owner and directors of the company, in order to obtain a balance of control and accountability.⁹⁸ Due to the fact that shareholders are liable to suffer the consequences of any wrong-doing carried out by the directors in their name, they should be authorised to issue any type of instructions to the directors concerning the performance of their duties.⁹⁹

Different mechanisms and principles established with the purpose of reducing agency cost problems are adopted in different legal systems.¹⁰⁰ In China, the very high concentration of ownership is directly linked to control from the board of directors, which is regarded as a critical connection between ownership and control in the current Chinese corporate governance scheme.¹⁰¹ Efforts have been made by the government to reduce the over-concentration of state shareholding. Notwithstanding a general objective to create a more dispersed and competitive shareholding structure, corporatisation and ownership diversification have also led to the emergence of new owners such as individual minority shareholders, institutional investors and employee shareholders.¹⁰² However, state ownership still prevails in most listed enterprises where the state still holds a high percentage of ownership.¹⁰³ These unique characteristics of Chinese corporate governance make derivative action a particularly important remedy for minority shareholders because of the difficulties of removing the directors or bringing about other changes.

Chinese corporate governance is at a primitive stage, and serious misbehaviour on the part of the board of directors is the major problem.¹⁰⁴ Deterrence by way of legal sanction is the solution, since voluntary market discipline would not function as efficiently in China as in countries with a mature market economy.¹⁰⁵ Therefore, minority shareholders in China are forced to rely on negative control through legal litigation as opposed to positive control by voluntary measures. If this negative control can be used efficiently by shareholders, it will have a positive effect in enhancing managerial efficiency and accountability.

No shareholder wants to be liable for uncertain or very high costs but receive only a pro rata share of the gains from success, while other shareholders free-ride on his or her efforts. In order to enable shareholders to use negative control to its fullest extent, reasonable litigation costs are one of the key requirements. It is claimed that the effectiveness of shareholder litigation as a governance mechanism is hampered by collective action

97 UK Companies Act 2006, s 168.

98 Ramsay (n 95) 16.

99 A Carver, 'Corporate Governance: Capitalism's Fellow Traveler' in F M Patfield (ed), *Perspectives on Company Law* vol 2 (Dordrecht: Kluwer 1997) 69, 74–75.

100 See E Glaeser and A Shleifer, 'The Rise of the Regulatory State' (2003) 61 *Journal of Economic Literature* 401.

101 L H Tan and J Y Wang, 'Modeling an Efficient Corporate Governance System for China's Listed State-Owned Enterprises: Issues and Challenges in a Transitional Economy' (2007) 7 *Journal of Corporate Law Studies* 143, 147.

102 S Tenev, C Zhang and L Brefort, *Corporate Governance and Enterprise Reform in China: Building the Institutions of Modern Market* (World Bank and the International Finance Corporation 2002) 2–3.

103 S Vob and Y W Xia, 'Corporate Governance of Listed Companies in China' (Paper presented at Track 8 of IFSAM VIIIth World Congress 2006, hosted by VHB, Berlin, Germany, September 2006) 110–16 <www.ctw-congress.de/ifsam/download/track_8/pap00750.pdf>.

104 J Zhao, 'The Regulation and Steering of Corporate Social Responsibility in China: Stories after the Enforcement of Chinese Company Law' (2011) 22 *International Commercial and Company Law Review* 400.

105 Z Zhang, 'Legal Deterrence: The Foundation of Corporate Governance: Evidence from China' (2007) 15 *Corporate Governance: An International Review* 741.

problems, because the cost of bringing a lawsuit, although less than the shareholders' aggregate gain, is typically greater than a shareholder's pro rata benefit as a claimant.¹⁰⁶ The problem of financial incentives to police management raises a principal–agent problem; the attorney's incentives need not coincide with the shareholders' interests. As far as litigation cost is concerned, fixed litigation fees will encourage shareholders to sue on behalf of the company in order to reduce the agency cost. The fixed fee will also help to reduce monitoring costs incurred by shareholders since, in the case of derivative action, shareholder coordination is not necessary and fixed litigation fees will make the shareholders' decision easier even without any coordination. It seems reasonable that 'the availability of this action economizes on costs that otherwise would be necessarily incurred if shareholders were required to take collective action'.¹⁰⁷

Corporate governance is a process, a monitoring device and a mechanism to minimise problems that may be brought about in the agent–principal–third-party relationship. However, in a derivative claim, a new agent–principal–third-party relationship can be observed. The shareholder who brings the claim replaces the directors as the agent of the company as the principal, acting in the collective interests of all the shareholders. If the shareholder who brings the action, the agent in the new agency relationship, brings a derivative action on behalf of the company which is the direct beneficiary, the company is logically required to indemnify the cost of the claim simply because the litigation is brought for the company in the name of the shareholder, as is the case in a disclosed agency relationship. It is the general rule that, where a person contracts as an agent for a principal, the contract is the contract of the principal, not of the agent.

In China, the minority shareholder's right and incentive to bring action on behalf of the company becomes more important because of the shareholding ownership structure and *guanxi* between directors and minority shareholders. A reasonable and encouraging cost system is important for a more accessible derivative action system. Agency theory can be used as a theoretical support for a more sensible fee arrangement, including fixed litigation costs and the adoption of indemnity orders.

4.2 PATH DEPENDENCE THEORY

Path dependence, a comparatively new theory originating in the 1980s, suggests that an outcome or decision is shaped in specific and systematic ways by the historical path leading to it, as well as by other factors within the socio-economic context.¹⁰⁸ While convergence theorists predict that countries, especially countries with weak legal systems, will adopt certain legal rules that have been demonstrably efficient in other jurisdictions, theorists who adhere to path dependence normally argue that divergence will still exist because legal rules are shaped by pre-existing political and social forces.¹⁰⁹ As part of the domestic legal and financial framework, a corporate law system has significant sources of path dependence, which include historical accidents as well as economic and political particulars of the

106 R Romano, 'The Shareholder Suit: Litigation Without Foundation?' (1991) 7 *Journal of Law Economics and Organization* 55, 55.

107 A Reisberg, *Derivative Actions and Corporate Governance Theory and Operation* (OUP 2007) 24.

108 O A Hathaway, 'The Course and Pattern of Legal Change in a Common Law System' (2001) 88 *Iowa Law Review* 101, 103–04.

109 K Pistor, *Patterns of Legal Changes: Shareholder and Creditor Rights in Transition Economies* (European Bank for Reconstruction and Development Working Paper 49 2000).

domestic system.¹¹⁰ The persistence of these sources significantly contributes to the stability of the domestic corporate governance system in any local socio-economic environment.

Path dependence theory can be regarded as a theoretical base for the adoption of a derivative action system with many unique characteristics shaping a particular nation's unique corporate governance model, corporate law background, enforcement process, shareholder structure, civil procedure law, and stage of economic development. It has been argued by Bebchuk and Roe that the initial ownership structure in a country will directly influence the subsequent development of ownership structure and laws.¹¹¹ Furthermore, these authors developed the theory to suggest that the interested parties possessing the power to influence ownership structure and corporate law will have both the incentive and the power to impede changes that might improve efficiency but are contrary to their private control interests.¹¹²

In China, the term 'listed company' normally refers to listed SOEs, and the largest shareholder normally refers to the state, which dominates the shareholding in Chinese listed companies. The very high concentration of state ownership is directly linked to control from the board of directors, whose members are regarded as a critical link between ownership and control in the current Chinese corporate governance scheme.¹¹³ However, when the SOEs are listed, the local office of the Bureau of State Asset Management or its local subsidiaries always act as the largest shareholders, and the chair of the board of directors is usually a representative from the Bureau who will consult with the board and nominate the directors of the SOEs.¹¹⁴ Therefore, the decisions made by the board will be closely focused on the largest shareholder's interests. This unique shareholding structure makes derivative lawsuits from minority shareholders particularly valuable while the prevalence of SOEs and the concentrated ownership structure renders minority shareholders extremely vulnerable.

Furthermore, the legal environment for the protection of minority shareholders is important in determining the strength of a country's capital markets.¹¹⁵ The stock market in China is under-developed and market forces are very weak.¹¹⁶ The reason for most shareholder investment in China is to gain immediate profits from dividends. Minority shareholders are always ignorant of company performance and information disclosed, and they own stock for a relatively short shareholding period.

The arguments concerning modifications of the standing requirement in joint stock limited companies seem to be reasonable and consistent with the Chinese corporate governance framework. The unique shareholding structure and the dominant position of directors and controlling shareholders make it necessary to adapt current company law to offer efficient remedies to minority shareholders and ensure effective supervision of directors. The speculative business virtue of minority shareholders' investment and their

110 L Bebchuk and M J Roe, 'A Theory of Path Dependence in Corporate Governance and Ownership' (2000) 52 *Stanford Law Review* 127.

111 Bebchuk and Roe (n 110).

112 *Ibid* 132.

113 Tan and Wang (n 101) 147.

114 J F Huchet and X Richer, 'China in Search of an Efficient Corporate Governance System: International Comparison and Lessons' (Centre for Economic Reform and Transformation, Herriot-Watt University, Edinburgh, Discussion Paper No 99/01, 11 February 1999).

115 R Tomasic and N Andrews, 'Minority Shareholder Protection in China's Top 100 Listed Companies' (2007) 9 *Australian Journal of Asian Law* 88; see also H Cai, 'Bonding, Law Enforcement and Corporate Governance in China' (2007) 13 *Stanford Journal of Law, Business and Finance* 82.

116 M Xu and S Neftci, *China's Financial Markets* (Elsevier 2007) 218.

remote distance from business operation in China make these reforms even more necessary. The rule-oriented protection of minority shareholders in the domain of corporate law shows its merits in comparison with market forces in China.

There is another argument based on path dependence theory that could explain the relationship between the unique corporate governance system in China and its derivative claim system. In terms of problems funding the litigation costs, a solution that has been advanced and practised elsewhere is the establishment of a foundation that processes a certain amount of the shares from each listed company in that jurisdiction.¹¹⁷ Corporate governance-related litigations funded by similar non-profit organisations have proved reasonably successful in Taiwan, Japan and South Korea.¹¹⁸ However, it is doubtful whether this system could be enforced in China because of the Chinese corporate governance and business organisation structures. The transplant of the system into China would be hindered by the government's tight control over civil society instructions.¹¹⁹

More generally speaking, China's listed companies face a set of very different ownership, business and financial conditions in comparison with companies operating under the Anglo-American model, and institutional conditions for the successful operation of the Chinese corporate governance framework are either absent or undeveloped.¹²⁰ Fiduciary duties of directors are not clearly addressed in CCL 2006 and issues of standards of liability for independent directors and directors in general remain unclear.¹²¹ The legal system of the People's Republic of China, with little more than half a century's history, is still in its infancy compared with many continental European or Anglo-American law systems. The division between the supervisory power of the Chinese Securities Regulatory Commission and the authority of stock exchanges is unclear, leading to confusion in the regulation of the securities market and corporations.¹²² The enforcement of Chinese law is regarded as one of the most challenging problems for the government, despite the high priority which has been placed on judicial development.

The enforcement of law and regulations for corporate governance will also be different in China than they are elsewhere. Little judicial independence can be observed even in modern China because the judiciary is 'parallel to rather than superior to, other units of the Chinese bureaucracy'.¹²³ The courts in China are 'people's courts', they are financed by government and are administratively and institutionally accountable to the government.¹²⁴ There is a lack of thorough understanding and sufficient experience on the part of

117 P Jiang, *The Development of Corporate Law from the 19th to the 20th Century (Gongsifa Cong 19 Shiji Dao 20 Shiji de Fazhan)* in F Guo and J Wang, *A Comprehensive Study of Company Law Reform (Gongsifa Xingai Zongheng Tan)* (Beijing: Law Press China 2000) 204.

118 C J Milhaupt, 'Nonprofit Organisation as Investor Protection: Economic Theory, and Evidence from East Asia' (2004) 29 *Yale Journal of International Law* 169.

119 Clarke (n 84) 55.

120 O K Tam, 'Ethical Issues in the Evolution of corporate governance in China' (2002) 37(3) *Journal of Business Ethics* 303, 311.

121 D C Clarke, 'The Independent Directors in Chinese Corporate Governance' (2006) *Delaware Journal of Corporate Law* 125, 224.

122 Wei (n 82) 485.

123 M A Layton, 'Is Private Securities Litigation Essential for the Development of China's Stock Market?' (2008) 83 *New York University Law Review* 1948, 1957–58.

124 *Ibid* 1958.

directors, judges,¹²⁵ solicitors and other practitioners regarding actual legal meanings, the nature and scope of their duties and how to enforce them through management policies. However, transplanting the derivative action system to China and making it enforceable will involve a long process based on a more solid and thorough understanding of the nature and scope of the directors' duties codified in CCL 2006, together with more expertise from judges and more in-depth academic research. Thus, whether shareholders' derivative action in CCL 2006 will truly enable shareholders to recover from mismanagement remains uncertain, and the accessibility of the system is particularly important in China due to the doubtful ability of the courts, as well as a lack of confidence in the courts on the part of minority shareholders and the public. China is the first Communist nation in the world to have a stock exchange, and the only socialist country to initiate the creation of a market-style modern enterprise system since 1990, through a corporatisation and shareholding framework but without privatising its SOEs. China needs a unique system of its own, shaped by the 'socialist market economy with Chinese characteristics'. This is also the case for a putative Chinese derivative action system.

5 Conclusion

Clark labelled the shareholder derivative suit as 'one of the most interesting and ingenious accountability mechanisms' for large companies.¹²⁶ The introduction of the shareholder derivative action system in CCL 2006 has been regarded as a milestone for CCL and corporate governance reform.¹²⁷ The new CCL 2006 offered many changes to various aspects of corporate governance and made it easier to start and conduct business by using corporations as a form of business organisation in China. The statutory derivative action framework enacted in China is, just like most of its predecessors, designed to correct the difficulties faced by minority shareholders who are faced with controllers who have misused their power within the company. With the dramatic increase in investments from foreign institutional investors in the emerging market, from US\$25 billion in 1990 to US\$300 billion in 2005,¹²⁸ the acceptance of shareholder derivative litigation demonstrates the way in which emerging economies like that in China are seeking to attract domestic and foreign investors and compete with developed economies for capital.¹²⁹ The new law also enhances the protection of disadvantaged groups such as minority shareholders and various stakeholders, addressing some of the issues arising from ineffective corporate governance systems and the abuse of power from boards and controlling shareholders. The derivative action system offers a legitimate platform to redress corporate misbehaviour in the interests of minority shareholders and the company at large.

125 With more than 170,000 law graduates every year gaining their LLB degree from universities, the qualifications of Chinese judges are worrying. Only 19.1% of them have a Bachelor's degree or higher, while many have not graduated from law school and a substantial number graduating do so through continuing education; see M Zhang, 'International Civil Litigation in China: A Practical Analysis of the Chinese Judicial System' (2002) 25(1) *Boston College International and Comparative Law Review* 59, 94–95.

126 R C Clark, *Corporate Law* (Aspen Publishers 1986) para 15.1.

127 N Krause and C Qin, 'An Overview of China's New Company Law' (2007) 28 *Company Lawyer* 316, 316, 319; see also M Gu, *Understanding Chinese Company Law* (2nd edn, Hong Kong UP 2010) ch 6.

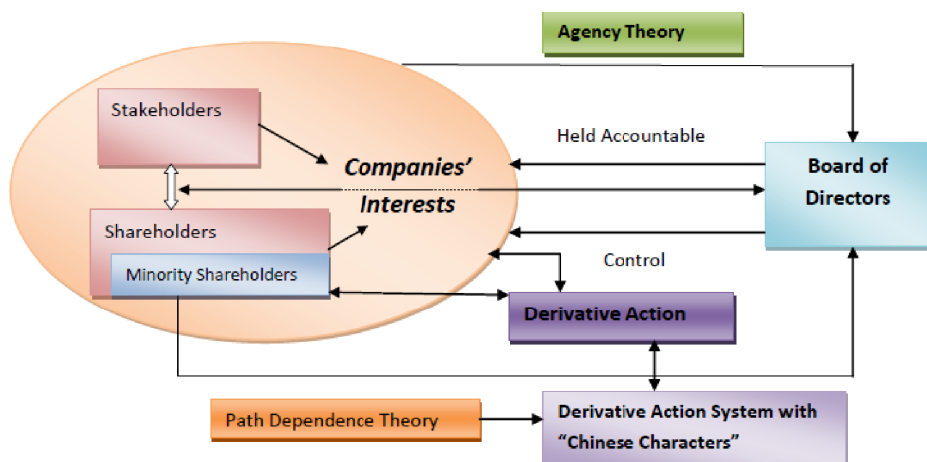
128 T Moss, V Ramachandran and S Standley, *Why Doesn't Africa Get More Equity Investment? Frontier Stock Market, Firm Size and Asset Allocation of Global Emerging Market Fund* (Centre for Global Development Working Paper No 112 2007) 1, <<http://ideas.repec.org/p/cgd/wpaper/112.html>>.

129 A M Scarlett, 'Investors Beware: Assessing Shareholder Derivative Litigation in India and China' (2011) 33 *University of Pennsylvania Journal of International Law* 173, 177.

Compared with pre-existing legislations, uncertainties about derivative suits in China¹³⁰ have been reduced by the amendments brought in by CCL 2006. However, in terms of the newly established derivative action, ineffective aspects within the new statutory system stipulated in CCL 2006 still hinder the enforcement of an investor-friendly derivative lawsuit system in China. Doctrinal and political obstacles severely limit the function and efficacy of the derivative suit as a device for policing management malfeasance.¹³¹ These deficiencies rest on standing, scope and the cost of litigation. Reform suggestions have been made regarding each deficiency of the system, with the purpose of building a new and efficient framework by maintaining a balance between granting rights to minority shareholders while at the same time avoiding frivolous suits. The suggestions were made clearer by discussions of theoretical support from corporate governance theories including agency theory and path dependence theory. It is concluded that derivative action in China must include a few specific local variations to fit with the many unique characteristics of China's history, economy and society. The relationship between the interests of shareholders, stakeholders, the company, directors, agency theory and stakeholder theory can be seen in the Figure 1.

Priority should be placed on encouraging shareholders to take derivative action and address the misconduct of company management teams, rather than on restricting the scope of derivative action.¹³² However, there is still a long way to go before China can fully adopt a derivative action litigation system with efficient Chinese civil procedures, thereby offering shareholders better and more practical remedies.

Figure 1: Corporate actors, corporate decisions and corporate governance theory



130 The uncertainties can be observed in statutory instruments and also in cases, for example: arts 111 and 63 of CCL 1993; art 42 of Chinese Securities Law 1999, art 4 of Principles of Corporate Governance for Listed Companies; *Zhangjiagang Fiber Co case* (1994); *Shanghai Yangzhong Water Company case* (1997); *Xiamen Xinda Network Co case* (1997); *Zhejiang Wu Fang Zhai Co case* (2001); and the more recent *Sanjin Pharmaceutical Co case* (2003).

131 Clarke (n 84) 47.

132 G Li and C Wang, 'Issues on Adjudicating Cases of Company Law Litigation: Stories after the Enforcement of Chinese Company Law 2006' available (in Chinese) <www.civillaw.com.cn/article/default.asp?id=46000>.

Beyond financialisation? How direct regulation will transform the future of financial market structures

DIETER PESENDORFER*

School of Law, Queen's University Belfast

1 Introduction

Over the last few years since the financial crash in 2008, it has become clear that the structure of financial markets is a major threat for financial stability and wealth, and this not just for countries as diverse as Austria (with its extensive exposure to Central and Southern Eastern European countries), Cyprus, Iceland, Sweden (with a banking crisis in the early 1990s), or the UK, where the financial sector has grown to a proportion of the gross domestic product that could only too easily bring down these economies entirely, not to mention the global effects that could occur in a worst case scenario. Throughout the most developed countries modern finance has become a too complex system with a huge variety of financial institutions which no longer only perform services of financial intermediation in terms of classical banking definitions. Massively driven by speculation, this system is characterised by a dominance of internationally active megabanks and large financial conglomerates; by complex national and transnational interconnections between banks as well as between banks and the 'shadow banking' world; by excessive leverage; and by a lack of transparency especially with regard to these opaque structures and the underlying products and activities that reproduce them. The global financial crisis itself, as well as numerous scandals, such as the 2010 Flash Crash, the spectacular JPMorgan's 'London whale trade losses' in 2012, or the Libor manipulations, illustrated to the public that this system and its risks cannot be managed adequately and that regulatory oversight remains a mission impossible. Although most megabanks have returned to large profits and so-called systemically important financial institutions (SIFIs) have passed official stress tests for which they had to increase capital buffers, problematic pre-crisis trends have continued and partly even intensified. With a *New York Times* bestseller book title describing the situation at the time of the crash, one could argue that 'all the devils'¹ are still around or side with Admati and Hellwig that 'even with proposed reforms' the emerging system 'is as dangerous and fragile as the system that brought us the recent crisis'.² Moreover, the too-big-to-fail (TBTf) institutions still profit in various ways from implicit government

* d.pesendorfer@qub.ac.uk.

1 B McLean and J Nocera, *All the Devils Are Here: Unmasking the Men Who Bankrupted the World* (Portfolio/Penguin Press 2010).

2 A Admati and M Hellwig, *The Bankers' New Clothes: What's Wrong with Banking and What to Do about It* (Princeton UP 2013) xii.

subsidies which are no longer seen as justifiable within an economic paradigm based on the idea of the superiority of highly competitive markets.

Emerging parallel with the overcomplex structures a cultural change in banking happened over recent decades that economists such as Paul Krugman or John Kay described as one from 'boring banking' of the 1950s and 1960s to a highly dynamic, innovative and fast-growing banking system, finally culminating in an excitement in this sector only comparable to the situation before the Wall Street Crash in 1929.³ The result is an emerging literature and political debate on the causes and effects of structures, cultures, incentives, products and activities. Many have analysed modern finance as 'casino banking' driven by recklessness, greed, corruption, and gambling,⁴ by 'mad money',⁵ 'madness with method',⁶ or 'medieval alchemy'⁷ which undermines the 'real economy' and increases overall inequality and instability. This system, labelled by critics as 'casino capitalism',⁸ 'financialization'⁹ and a 'finance-led accumulation regime',¹⁰ has produced numerous national and regional crises since its emergence in the 1970s, before finally proving how unstable, risky, destructive and costly it has become globally. However, before the global financial crisis, mainstream actors were convinced that modern finance had become safer thanks to deregulation, liberalisation, 'light-touch' regulation, business self-regulation and sophisticated risk management practices, such as diversification via securitisation. Stricter monetary policy, the trend towards independent central banks, and further policies described by critics as neoliberal were seen as major improvements that had led to a 'Great Moderation' with less macro-economic volatility allowing more leverage and high liquidity.¹¹ Emerging markets and poor countries were seen as remaining sources for future financial crises and reform proposals focused on measures that would make their regulatory systems similar to those in the most developed countries.¹²

Now with the continuing global financial crisis in its various stages and multiple facets, its aftershocks and the return of 'crisis economics'¹³ with all kind of uncertainties and pressures to reform the very role of modern finance, its contemporary size, shape and dynamics are radically put into question. Proponents of the shareholder value paradigm propose radical reforms to rescue this idea from its distortions caused by short-termism.¹⁴ Economists as an entire discipline have started to explain how it came to the enormous

3 P Krugman, 'Making Banking Boring', *New York Times* (New York, 10 April 2009); P Krugman, 'Good and Boring', *New York Times* (New York 1 February 2010); J Kay, *Narrow Banking: The Reform of Banking Regulation* (2009) <www.johnkay.com/wp-content/uploads/2009/12/JK-Narrow-Banking.pdf>.

4 G Morgenson and J Rosner, *Reckless Endangerment: How Outsized Ambition, Greed, and Corruption Led to Economic Armageddon* (Times Books 2011).

5 S Strange, *Mad Money: When Markets Outgrow Governments* (University of Michigan Press 1998).

6 S Wagenknecht, *Wahnsinn mit Methode: Finanzcrash und Weltwirtschaft* (Das Neue Berlin 2009).

7 K Dowd and M Hutchinson, *The Alchemists of Loss: How Modern Finance and Government Intervention Crashed the Financial System* (John Wiley & Sons 2010); A Nesvetailova, *Financial Alchemy in Crisis* (Pluto Press 2010).

8 S Strange, *Casino Capitalism* (Blackwell 1986).

9 G A Epstein (ed), *Financialization and the World Economy* (Edward Elgar Publishing 2005); G Krippner, *What is Financialization?* (mimeo, Department of Sociology, UCLA 2004).

10 M Aglietta, 'Capitalism at the Turn of the Century: Regulation Theory and the Challenge of Social Change' [1998] (Nov/Dec) NLR 41.

11 B S Bernanke, *The Great Moderation* (remarks by Governor Ben S Bernanke at the meetings of the Eastern Economic Association, Washington, DC, 20 February 2004) (Federal Reserve Board 2004) <www.federalreserve.gov/boarddocs/speeches/2004/20040220/default.htm>.

12 B Eichengreen, *Financial Crises and What to Do About Them* (OUP 2002).

13 N Roubini and S Mihm, *Crisis Economics: A Crash Course in the Future of Finance* (Allen Lane 2010).

14 A Rappaport, *Saving Capitalism from Short-Termism: How to Build Long-Term Value and Take Back Our Financial Future* (McGraw-Hill 2011).

'misunderstanding' of financial crises¹⁵ but still struggle to understand what lessons to be drawn.¹⁶ System transformation with deleveraging and many other measures affecting the future structures of finance came on the agenda. Asking about the benefits of the size, structure and functions of finance and demanding quite radical change, as well as thinking about the optimum number and size of banks, about ways to decrease interconnectedness among firms and to improve competition, and about the overall future structures of finance have become not just popular among many regulators and law-makers but a key policy goal already resulting in numerous regulatory changes and proposals.

This paper analyses the various reforms and proposals and argues that in the area of structural reforms reformulations still remain within the overall system of financialisation. The adopted reforms include a number of significant loopholes and flaws that altogether lead to a result in which the overall dominance of finance will persist and create continued financial instability. By making regulatory proposals overly complex, loopholes and flaws were increased. Additionally, institutional arrangements remain problematic. Significant problems and uncertainties with regard to overly complex and opaque banking structures remain. However, the overall direction of reforms might still change in key jurisdictions and heavy aftershocks might quickly bring more radical, yet marginalised, proposals on the agenda. At least in theory one could assume that several reforms potentially lead to a more significant change via gradual institutional reforms, while more radical reforms would require much more political pressure and will. The next section will discuss why and how structure matters and why financial structures became a problem that needs to be resolved. Section 3 will look at various regulatory reforms and proposals addressing financial structures with a focus on measures that affect financial institutions directly, such as most prominently the Volcker Rule and ring-fencing. The concluding section will problematise the regulatory space for taming finance in times of crisis, austerity and increased public protest potential. It will be concluded that new regulatory strategies are required that draw the lessons from the failed radical reform proposals so far.

2 Structures of finance matter

Half a decade after the meltdown, regulators in a still highly politicised policy area are facing great expectations and public pressure to address the structural flaws of the current financial system and to make the future structures more robust and resilient. Before the crisis regulating finance was basically a rather closed expert discourse with moderate and radical critics marginalised. The dominant paradigm shared by business and financial markets experts and regulators included the assumption that the general trends towards higher structural complexity within the liberalised and deregulated markets were extremely positive and allowed unprecedented economies of scale and scope, increasing transnational activities, extreme leverage, and a focus on investment banking which would result in overall benefits for societies. That the ever larger financial institutions resulting from fast capital concentration through global competition, mergers and acquisitions would be more stable, because such institutions would be more diversified and superior in applying new risk measurement and management techniques, was another credo of the time.¹⁷ Within the European Union (EU) the demand for a highly integrated financial internal market with several global players was another key driver for ever more complex structures. Financial centres such as Wall Street or the City of London, as well as offshore centres, were highly respected and seen as positive drivers of globalisation contributing to more growth and

15 G B Gorton, *Misunderstanding Financial Crises: Why We Don't See Them Coming* (OUP 2012).

16 E Porter, 'Economists Agree: Solutions Are Elusive', *New York Times* (New York, 23 April 2013).

17 M Carey and R M Stulz (eds), *The Risks of Financial Institutions* (University of Chicago Press 2006).

wealth across the planet. Financialisation – understood as the increased importance of financial motives, markets, actors and institutions – would allow all social groups and every individual to participate in this new world of better and more democratic access to financial products. The overall pre-crisis paradigm ignored systemic problems caused by the increasing complexity of structures, activities and products, extreme leverage, the increasing interconnectedness and size of financial institutions and reduced transparency. In particular, calls for a global systemic risk regulator receiving broader attention in the aftermaths of the East-Asian financial crisis of 1997–98 were ignored. Largely overlooked were also problems of unfair competition derived from subsidies for TBTF institutions, or the fact that various fast-growing markets, such as high-frequency trading or derivatives trading, were highly concentrated giving unfair advantages to specific market players, such as megabanks, high-frequency traders or the privileged ‘primary dealers’. It is only in the post-meltdown climate that ever larger, more complex and more leveraged financial institutions and unregulated, market-driven innovation are no longer uncritically accepted by key regulators but seen as unacceptable causes for excessive risk-taking, excessive leverage, and inadequate capital requirements that altogether increase the probability of crises and bailouts.

Given poor evidence, it is rather odd how complex and opaque structures, products and activities and, in particular, high leverage and unregulated innovation are often presented as absolutely superior and essential for global markets. In 2009, Paul Volcker, a former chair of the US Federal Reserve and an outspoken critic of today’s financial system, polemicised against this ideology-driven claim that the only useful innovation of modern finance was the automated teller machine (ATM) and that he would still await someone showing him evidence that modern finance would actually contribute to economic growth.¹⁸ In the UK, Adair Turner, then chair of the now dismantled Financial Services Authority (FSA), similarly concluded:

There is no clear evidence that the growth in the scale and complexity of the financial system in the rich developed world over the last twenty to thirty years has driven increased growth or stability, and it is possible for financial activity to extract rents from the real economy rather than to deliver economic value.¹⁹

Martin Wolf, chief economics commentator of the *Financial Times*, repeatedly argued that the crisis resulted from ‘an illmanaged, irresponsible, highly concentrated and undercapitalised financial sector’ and that we have to be really radical this time as the crisis so far brought us into a situation in which ‘the most important point is that where we are now is intolerable’.²⁰ Similarly, Sir Mervyn King, Governor of the Bank of England until mid-2013, stated in a speech:

Of all the many ways of organizing banking, the worst is the one we have today . . . Change is . . . inevitable. The question is only whether we can think our way through to a better outcome before the next generation is damaged by a future and bigger crisis.²¹

18 ‘Volcker Praises the ATM, Blasts Finance Execs, Experts’, *Wall Street Journal* (New York, 8 December 2009); ‘Paul Volcker: Think More Boldly’, *Wall Street Journal* (New York, 14 December 2009).

19 A Turner, ‘What Do Banks Do? Why Do Credit Booms and Busts Occur? What Can Public Policy Do About It?’ in A Turner et al (eds), *The Future of Finance: The LSE Report* (London School of Economics 2010).

20 M Wolf, ‘Narrow Banking Alone is Not the Answer’, *Financial Times* (London, 15 December 2009).

21 M King, *Banking: From Bagehot to Basel, and Back Again* (Second Bagehot Lecture Buttonwood Gathering, New York City, 25 October 2010) (Bank of England 2010) <www.bankofengland.co.uk/publications/Documents/speeches/2010/speech455.pdf>.

In short, ‘international policymakers have concluded that many of the world’s largest banks are not the right size’ and are now looking at ways to make banking structures more ‘robust and resilient’.²²

One of the core questions in contemporary financial regulation reform debates is about the best – the optimum – structure for traditional banking to reduce systemic risk and to stop TBTF by removing any implicit or explicit government support for financial institutions (including the shadow banking system) that creates moral hazard and the necessity of bailouts by the taxpayer. In a functioning market economy, failing firms are expected to go into bankruptcy, which in turn improves the situation for the surviving businesses – this is why Joseph Schumpeter²³ described this process as ‘creative destruction’ and a key driver of economic development. But TBTF seemed to make bailouts, state aid and mergers a necessity that generally led to further market concentration and market distortions. These trends have continued over the past few years. Moreover, an increase in mergers is already expected with market consolidation in a calmer post-crisis period. In the crisis situation, subsidies create the problem of the so-called zombie banks and various restrictions on legal action (‘too big to jail’) which not only undermine the functioning of the market but also the legitimacy of the capitalist system. Even by neoliberals these fundamental market distortions caused by implicit or explicit government support are understood as in urgent need of reform.

Deleveraging and less complex, more competitive and more transparent structures are regarded as necessary to reduce systemic risk and to design a more preventive and precautionary financial system.²⁴ Yet, still only a minority wants to return to boring banking, or an even more radical overhaul of the financial structures. Ideas to redesign finance along the lines of strict product and activities regulations, such as – despite several flaws – is already the practice in drugs, food, or chemicals regulation in various regimes, did not find broad support either. Proposals, such as learning from Islamic banking, largely unaffected by the global crisis, were also only supported by a few. The majority of regulators and especially of business representatives still hope to get away with rather moderate adjustments. A too simple and restrictive system is by many seen as no longer adequate given global trade levels and interconnectedness and is often associated with a loss in overall consumer welfare. Hopes that this crisis could transform the system beyond financialisation are counteracted by significant fears that any such an attempt would go in a wrong direction or even that modest reforms would already go too far by creating more harms than goods, especially in this time of economic recovery in which finance is needed.

The question about the optimum structure is related to the insight that the industry structure has to be transformed by regulation and not just by post-crisis market or technology-driven processes that modify the structures of finance too. Changing industry structures via regulation and the right timing of this approach are, of course, highly contested. It is only logical that industry opposes any changes to a system that generates high returns and bonuses. Over recent years, there were numerous complaints from business lobbyists that policy-makers would wrongly follow a populist demand for

22 A G Haldane, *On Being the Right Size* (Bank of England, speech, Institute of Economic Affairs, Pall Mall, 25 October 2012) <www.bankofengland.co.uk/publications/Pages/speeches/default.aspx>.

23 J A Schumpeter, *Capitalism, Socialism and Democracy* (Routledge 1992).

24 J Crotty and G Epstein, ‘A Financial Precautionary Principle: New Rules for Financial Product Safety’ (Wall Street Watch Working Paper, No 1, Consumer Education Foundation 2009); D Pesendorfer, ‘Goodbye Neoliberalism? Contested Policy Responses to Uncertain Consequences of the 2007–09 Financial Crisis’ in K Alexander and R Dhumale (eds), *Research Handbook on International Financial Regulation* (Edward Elgar Publishing 2012) 414–34.

punishing bankers resulting in over-regulation with disastrous effects on the financial sector's contributions to economic recovery. Further concerns are that, when not in line with international agreements, too ambitious reforms might just lead to economic disadvantages caused by regulatory competition on globalised financial markets and that policy-makers and regulators are limited in their actions by fears that institutions and activities would just move to other jurisdictions in response to tough regulation. Moreover, modest reforms are supported by contemporary regulatory practices. Opponents of more significant structural reforms frequently point towards high costs in such processes and uncertainties of measures, such as much stricter capital requirements, breaking up the largest institutions, controlling innovation, or more transparency. Good governance regulatory tools such as public consultations and Regulatory Impact Assessment with Benefit-Cost Analysis are only of a limited value in these discussions. Public consultations are dominated by the opinions of experts from the financial sector beside the various other 'behind-the-door' lobbying activities.²⁵ Regulators are also often captured by business interests.²⁶ It is moreover extremely hard to calculate any possible effects of structural reforms over the next few decades and all proposals necessarily include a high level of uncertainty.²⁷ Frequently powerful lobbying groups intervened with calculations of the high costs of reforms, while ignoring the significant costs of the likely next crash and crisis.²⁸ In short, we not only have insufficient evidence about the benefits and costs of the current financial system but also about the various options for its long-term transformation, which leaves decision-makers with a political decision about what the future of finance should look like and a normative decision about how much stability this will create and what risk-risk trade-offs are acceptable. And up to now, the political will for real reforms was mostly missing, as Admati and Hellwig lamented.²⁹ In this climate, some observers have concluded that there has been little real change so far. Others have started to assume that the major decisions to shape the structures of finance for the next few decades have been made. For the US, Simon Johnson concluded: 'The debate is over; the decision to cap the size of the largest banks has been made. All that remains is to work out the details.'³⁰ What kind of decisions have been made and in what direction the reforms will go is the focus of the next section.

3 Redesigning and transforming finance

Structures of finance can be transformed directly or indirectly and either by building on existing regulation or by adopting new regulation. Indirect regulation affects the overall market dynamics by correcting incentives for specific products and activities that then affect market structures. Such price regulations are, for example, higher capital requirements or sector-specific taxes. Direct measures aim at regulating structures of financial institutions via institution-specific requirements. From among the numerous reforms of finance adopted or proposed over the past few years, this paper will focus on proposals to change the structure of finance directly via structural regulation. Given the adopted measures and proposals so far, divergence between the US and the EU, but even within EU member

25 Admati and Hellwig (n 2).

26 S Pagliari (ed), *Making Good Financial Regulation: Towards a Policy Response to Regulatory Capture* (Grosvenor House 2012).

27 See eg the Impact Assessment for the UK Financial Services Bill 2012.

28 S Johnson, 'The Impact of Higher Capital Requirements for Banks', *New York Times Economic* (New York, 18 April 2013); Oxford Economics, *Analyzing the Impact of Bank Capital and Liquidity Regulations on US Economic Growth* (Clearing House and Oxford Economics 2013).

29 Admati and Hellwig (n 2).

30 S Johnson, 'The Debate on Bank Size Is Over', *New York Times* (New York) 28 March 2013.

states, has become likely. However, numerous loopholes and flaws in regulations might make the differences less important as will be argued. Moreover, there have always been quite significant dissimilarities in the financial structures of these jurisdictions and some differences clearly accommodate the chosen regulatory paths. The most important ideas with regard to becoming in some form regulatory reality are the US approach linked to the Volcker Rule and the British/European reform proposal of ring-fencing, as well as reforms on orderly resolution. But there are also a number of other, more or less marginalised demands, such as opposition to interventions, ‘narrow banking’, limited purpose banking, or even further nationalisations of banks, that need to be explored given the dynamics of a crisis that might allow ‘radical’ proposals to become dominant. This section will discuss these ideas with regard to their overall effects including those related to regulatory competition, intended and unintended consequences, possible weaknesses, and flaws and loopholes in order to ask whether any of these proposals is radical enough to make the financial system safer and to achieve the stated goals.

3.1 NO REGULATION OF INDUSTRY STRUCTURES

One of the major claims in this discourse is that in a situation of high uncertainty any new regulation would generate highly negative unintended consequences. Most neoliberals share the belief that governments should not intervene and try to change or design market structures. This should be solely left to market forces, whereby one market-based solution to deal with systemic risk might be to force firms to buy insurance against its potential losses.³¹ They also argue against bailouts or more interventionist merger control and many other measures that would affect the structures of finance. In the current historical situation, without a quick recovery and a return to pre-crisis conditions, the non-intervention demand is damned to remain ideology as neither firms nor regulators were highly supportive of radical market-based solutions. The argument with regard to structural reforms has therefore slightly shifted to emphasising that banking regulation has already been strongly strengthened via Basel III (still in need of fine-tuning and implementation) and that market-driven and corporate governance reforms have already brought significant changes in behaviour that all together will make banks safe enough. Structural reforms are now ‘untimely’ and would only harm economic recovery. Moreover, many bankers have claimed that structural reforms always pose assessment problems and would necessarily lead to negative unintended consequences. In the UK, they have emphasised that it would not have been the largest institutions that got into trouble and mega-size would not automatically translate into excessive risk. Similarly, in the US JPMorgan Chase was an often-used example of a big institution getting through the crisis without serious problems and its chief executive officer Jamie Dimon became the most prominent opponent against reform calls. However, how well such institutions performed during the crisis is not relevant for the question of how they contribute to a too high systemic risk level. Also regulators are looking at this issue and have concluded that size of the balance sheet alone is an insufficient indicator to capture problems related to more complex interconnections.³²

In the US, Mitt Romney’s defeat in the 2012 presidential elections ensured that structural reforms adopted with the Dodd-Frank Act 2010 will not be achieved for at least another four years. However, Republicans were successful in blocking proposals that regulators get sufficient funding for implementing reforms and worked actively towards delaying

31 V V Acharya et al, ‘Regulating Systemic Risk’ in V V Acharya and M Richardson (eds), *Restoring Financial Stability: How to Repair a Failed System* (John Wiley & Sons 2009) 283–303.

32 Eg the European Central Bank for this purpose developed a framework for identifying ‘large and complex banking groups’.

implementation. In the EU, where reforms have been slowed down because of the Eurozone crisis, the general acceptance for structural reforms is not necessarily higher with the UK highly concerned about the overall future competitiveness of London and with Germany and France also interested in improving their competitive position in global finance. With the European crisis, market-driven developments led to decreased transnational activities of banks and to problems meeting stricter capital requirements. Nevertheless, all the above-mentioned countries have put forward direct regulations affecting the future structures of finance. For Europe one could, of course, argue that a non-intervention position has already been strengthened over the past years with the 'modernisation' of European competition law and policy that brought the European approach closer to the Chicago School theory and away from interventions into markets in favour of deepening integration. Competition law is powerful and in its neoliberal interpretation on both sides of the Atlantic superior to other public policy goals. In the EU the unique state aid policy assures, moreover, that any state aid requires strict restructuring obligations.³³ Beyond the rather limited state aid policies, the framework largely favours non-intervention into the market mechanism. But, for now, the US is implementing the Volcker Rule, incorporated into the Dodd-Frank Act, and the EU Commission and the UK proposed ring-fencing as most important structural reforms, whereby the EU approach is more similar to the Volcker Rule and has recently been adopted by Germany in a new Banking Separation Act.³⁴

3.2 FROM GLASS-STEAGALL TO THE VOLCKER RULE

In the early 1930s climate of political and economic turmoil in the US, a debate had started about separating commercial and investment banking. President Hoover and Senator Carter Glass became advocates of this idea, sharing the analysis that this would solve problems that led to the stock market crash by eliminating various conflicts of interests and high-risk speculative lending practices. Representative Henry B Steagall was an advocate of deposit insurance to make banks safer. The reform proposals were finally also accepted by investment and commercial banks that saw at least temporary advantages in the final proposals.³⁵ The Banking Act of 1933, often referred to as the Glass-Steagall Act, was then the law establishing federal deposit insurance and various restrictions on speculation and splitting investment and commercial banking. Glass-Steagall's original intention was to bring all commercial banks under the Federal Reserve System and to establish strict restrictions on speculation. However, thanks to loopholes, a large number of state chartered banks remained outside this system, putting them also outside the restrictions on purchasing and selling securities. From the 1960s, the restrictions were already being interpreted less strictly and commercial banks were more and more allowed to engage in securities activities and, since the 1980s, investment banks have increased their engagement in thrifts or state chartered banks, avoiding Glass-Steagall affiliation restrictions and holding company laws. It nevertheless took until 1999, when under President Bill Clinton the Gramm-Leach-Bliley Act was adopted that removed the remaining barriers for commercial banks to engage in investment banking or to merge with insurance companies. The Glass-Steagall provisions against conflicts of interest were no longer seen as adequate regulation in modern financial markets but as unnecessary restrictions reducing market efficiency and liquidity and putting

33 U Soltész and C Von Köckritz, 'From State Aid Control to the Regulation of the European Banking System: DG Comp and the Restructuring of Banks' (2010) 6(1) *European Competition Journal* 285.

34 Gesetz zur Abschirmung von Risiken und zur Planung der Sanierung und Abwicklung von Kreditinstituten und Finanzgruppen, 17 May 2013.

35 V V Acharya et al, 'A Bird's-Eye View: The Financial Crisis of 2007–2009' in Acharya and Richardson (n 31).

US banks into competitive disadvantage. However, there were also early warnings that Gramm-Leach-Bliley would go too far and increase the TBTF problem.³⁶ But it was only in the aftermath of the financial meltdown that claims became widely popular that repealing Glass-Steagall was a major cause of the financial crisis and that commercial and investment banking should be split again. Only a few have argued for reintroducing Glass-Steagall and the dominant approach became a moderate reformulation under the label ‘Volcker Rule’.

The Volcker Rule, named after Paul A Volcker, a former Federal Reserve chair from 1979 to 1987 who championed the idea of breaking up the largest banks and prohibiting deposit-taking institutions from engaging in certain forms of risky activities, will restrict banks from speculating with federally insured deposits. More specifically, the rule prohibits bank holding companies from owning, investing in or sponsoring hedge funds or private equity funds and from engaging in proprietary trading. Volcker regards this measure as key ‘in the interest of fair and open competition as well as protecting the provision of essential financial services’.³⁷ This rule became a prominent cornerstone of US regulatory debates. President Obama proposed it in January 2010 as an integral part of the financial regulatory overhaul.³⁸ A House–Senate conference committee then approved the rule in June 2010.³⁹

Originally, Volcker outlined the rule ‘in a three-page letter to the president’. In 2010, when the Dodd-Frank Act ‘went to Congress, the Volcker Rule that it contained took up 10 pages’; in its public consultation version in October 2011 it was 298 pages, ‘accompanied by more than 1,300 questions about 400 topics’.⁴⁰

The Dodd-Frank Act sets out the Volcker Rule in s 619. The proprietary trading provisions in the Act prohibit:

a banking entity or nonbank financial company supervised by the Board’ from trading activity in which it acts as a principal (principal investments) for its own trading account ‘in any transaction to purchase or sell, or otherwise acquire or dispose of, any security, any derivative, any contract of sale of a commodity for future delivery, any option on any such security, derivative or contract, or any other security or financial instrument that the appropriate federal banking agencies, the Securities and Exchange Commission (SEC), and the Commodity Futures Trading Commission (CFTC) may, by rule as provided in subsection (b)(2), determine.

The hedge fund and private equity fund provisions generally prohibit deposit-taking institutions to ‘acquire or retain any equity, partnership, or other ownership interest in or sponsor a hedge fund or a private equity fund’.⁴¹

The Volcker Rule version in the Dodd-Frank Act was watered down compared to earlier formulations and included several exemptions. Although the rule bans banks from short-term trading of securities for their own account and limits investments in private equity groups and hedge funds, the restrictions allow a number of exemptions that have been defined as ‘permitted activities’ and longer-term commitments or ‘principal investments’. These activities include ‘hedging’, ‘market-making’, asset management, underwriting, and

36 G Stern and R Feldman, *Too Big to Fail: The Hazards of Bank Bailouts* (Brookings Institution Press 2009).

37 P Volcker, *Prohibiting Certain High-Risk Investment Activities by Banks and Bank Holding Companies* (Testimony before the Senate Committee on Banking, Housing and Urban Affairs, Washington DC, 2 February 2010).

38 *New York Times* (New York, 23 January 2010).

39 *New York Times* (New York, 26 June 2010).

40 *New York Times* (New York, 21 October 2011)

41 Dodd-Frank Act, SEC 619, s 13(a)(1)(B).

other transactions on behalf of customers. The line between these activities and purely speculative activities is highly flawed. For this reason Dodd-Frank is sometimes also labelled 'Glass-Steagall lite'.

The Dodd-Frank Act demanded the Financial Stability Oversight Council (FSOC) to study the implementation of the Volcker Rule and to make corresponding recommendations 'not later than 6 months after the date of enactment'. The FSOC published its report in September 2011.⁴² All the details of the rule's implementation were left by Dodd-Frank to the regulatory authorities (SEC, CFTC, Office of the Comptroller of the Currency, Federal Deposit Insurance Corporation, and the Board of Governors of the Federal Reserve System). The agencies would have been required to adopt implementation rules not later than nine months after the Financial Stability Oversight Council's study's completion. However, decision-making was slowed down by massive lobbying interventions and highly technical consultations. Critics insisting on a strict implementation were more and more marginalised between 2010 and 2012. The Republicans openly demanded dropping the Volcker Rule completely as part of their demand for reversing Dodd-Frank's over-regulation. Republicans argued in particular that the Volcker Rule would put US banks into clear competitive disadvantages on the global market.⁴³ They fought against a strict implementation in the US congressional bank oversight panel and by threats to block or reduce budgets for the agencies that will have to implement the Volcker Rule.

The final version was still delayed when, in May 2012, JPMorgan Chase had to disclose a \$2 billion trading loss (later turning out to be an at least \$6.2 billion loss)⁴⁴ that was broadly interpreted as strong evidence for the lack of change at Wall Street banks and their ongoing gambling that could only give advocates of a stricter implementation of the Volcker Rule a massive boost. In particular JPMorgan's loss led to debates about how this rule would have had an impact on such deals, had it already been in force, although JPMorgan did not release any details behind the trade. However, the flawed interpretations of hedging and market-making still allowing high-risk speculation and the consequences from weakening implementation gained attention. The FSOC had already emphasised the importance of reviewing 'permitted activities' closely and to implement a 'programmatic compliance regime' that would need to be 'designed to ensure that proprietary trading does not migrate into permitted activities'.⁴⁵ But with the JPMorgan loss scandal, demonstrating poor risk management in the top bank and the success of the large banks in interpreting 'hedging' and 'market-making' to their advantage, stricter implementation became at least for the moment more likely. The timing of this debacle was perfect for advocates as the final version of the Volcker Rule was expected by the summer or at least by early fall. This was an example of how business power, regulatory capture and regulatory competition pressures can become weaker under conditions of crisis, when the capacity for gradual institutional change exists. However, it is also an example of how the financial regulation elite still do not bother about huge profits coming from the same risky activities, but are surprised if things go wrong.

Compared to Glass-Steagall, the Volcker Rule is overly detailed and complex. Luigi Zingales became one of the few advocates for Glass-Steagall arguing that it is superior thanks to its simplicity and necessary for 'a well regulated and transparent public market':

42 FSOC, *Study and Recommendations on Prohibitions on Proprietary Trading and Certain Relationships with Hedge Funds and Private Equity Funds* (FSOC 2011).

43 Reuters, 4 November 2010.

44 Senate Permanent Subcommittee on Investigations, 'JPMorgan Chase Whale Trades: A Case History of Derivatives Risks and Abuses' (Permanent Subcommittee on Investigations, 15 March 2013).

45 FSOC (n 42).

The Glass-Steagall Act was just 37 pages long. The so-called Volcker rule has been transformed into 298 pages of mumbo jumbo, which will require armies of lawyers to interpret. The simpler a rule is, the fewer provisions there are and the less it costs to enforce them. The simpler it is, the easier it is for voters to understand and voice their opinions accordingly. Finally, the simpler it is, the more difficult it is for someone with vested interests to get away with distorting some obscure facet.

In Zingales' view, 'Glass-Steagall helped restrain the political power of banks' and was an important law against 'excessive power'.⁴⁶ With the Glass-Steagall, provisions investment banks, commercial banks and insurance companies would have different interests and agendas that would limit their overall power by off-setting one another.

Overall the highly controversial US debate reflects the huge uncertainties with regard to direct regulations that will shape the future of finance and that the current policy is not going beyond financialisation.

3.3 RING-FENCING

While the Volcker Rule is very much embedded in the US regulatory culture and memory of Glass-Steagall and in a strong antitrust policy concern against TBTF institutions, the idea gained much less support in Europe, where many countries have no history of a separation between retail and investment banking but extensive experience with universal banking. Inspired by a study by the Organization for Economic Co-operation and Development,⁴⁷ this led to proposals focusing on ring-fencing certain business areas by the UK Independent Commission on Banking (ICB) and by the European Commission's High-level Expert Group on reforming the structure of the EU banking sector.

In the UK, the Volcker Rule has mostly been criticised as the wrong approach. The largest banks argued that the crisis would not show a clear relationship between bank size and activities but only a relationship between good and bad risk management practices. Deposit-taking institutions engaged in proprietary trading and investing in hedge funds and private equity groups would provide useful services. The governments supported this view. Before forming a Coalition government, many Conservatives and Liberals were more critical about banking structures. George Osborne, then Shadow Chancellor, called for a sector with smaller banks. While former Prime Minister Gordon Brown rejected a 'rigid divide' of banks along a Glass-Steagall approach, Osborne argued that British banks had become TBTF and too big to be bailed out.⁴⁸ However, the Conservatives did not want to position themselves too early with precise plans. Their manifesto for the 2010 general election announced that they would 'pursue international agreement to prevent retail banks from engaging in activities, such as large-scale proprietary trading, that put the stability of the system at risk'. The Liberal Party, especially Vince Cable, called for a total separation of the retail banks from 'casino' investment banks. The coalition agreement between the Conservatives and Liberals, released on 12 May 2010, promised 'detailed proposals to foster diversity, promote mutuals and create a more competitive banking industry'. In order 'to reduce systemic risk in the banking system' they announced that they would 'establish an independent commission to investigate the complex issue of separating retail and investment banking in a sustainable way' within one year. On 16 June 2010, Chancellor Osborne announced the creation of this ICB chaired by Sir John Vickers.

46 L. Zingales. 'Why I Was Won Over by the Merits of Glass-Steagall', *Financial Times* (London, 11 June 2012).

47 A. Blundell-Wignall, G. Wehinger and P. Slovik 'The Elephant in the Room: The Need to Deal with What Banks Do', *Financial Market Trends* (OECD 2009).

48 *Financial Times* (London, 24 February 2009).

The ICB was a rather small commission, with two out of five members having had past or continued clear affiliation with financial firms. After taking up work, the ICB released an 'issue paper: call for evidence' in September 2010 that discussed a number of issues such as financial stability, competition, economic recovery and lending, competitiveness, and risks to the government's fiscal position, and presented reform options for consultation. The latter included options related to the structure of banks. These included the 'separation of retail and investment banking', à la Glass-Steagall, 'narrow banking and limited purpose banking', 'limits on proprietary trading and investing' in line with the Dodd-Frank Act's Volcker Rule, the degree of separation, structural separability, contingent capital, and structure-related surcharges. With regard to the structure of financial markets the ICB considered 'measures to reduce concentration' related to merger law and a market infrastructure reform with the goal of increased transparency. Against a Volcker Rule type of limits on proprietary trading and investing, the ICB argued that:

there is the practical question of how to tell the difference between proprietary trading on the one hand, and risk-reducing hedging or market-making on the other. For example, a bank might acquire a large position in interest rate swaps for speculative reasons or, by contrast, to hedge interest rate risk naturally arising from the maturity transformation of ordinary commercial banking business.⁴⁹

The ICB then accepted the arguments against a break-up of TBTF banks and proposed alternatively the idea of ring-fencing to protect retail banking from the investment banking divisions.⁵⁰ Banks would have to put their high-street banking in a separate legal subsidiary with higher capital requirements and an independent board. Instead of a total separation of retail and investment banking: both subsidiaries could remain under a common parent holding company. Banks would have some flexibility to decide what is inside and outside the higher protected retail part. The proposals should be implemented by 2019.

In total, the ICB believes that: 'Ring-fenced banks would be simpler, less connected and easier to resolve without taxpayer support than existing universal banks.'⁵¹

While the UK government supported the idea, ring-fencing has also been heavily criticised. The director-general of the Confederation of British Industry called it, in a *Financial Times* interview, 'barking mad'.⁵² Industry and lobbyists have argued that ring-fencing would lead to banks becoming uncompetitive and harm the chances of a recovery. Others called for high flexibility to ensure that the UK banks would not become uncompetitive against global competitors or argued that Basel III rules would be much more significant.⁵³ There was also some concern in the financial press over whether ring-fencing could work in practice.

After the Queen's Speech in May 2012 confirmed that the government would go ahead with the controversial measures within a Banking Reform Bill,⁵⁴ banks warned that plans to ring-fence would be too expensive and the costs would be passed on to businesses and household customers.⁵⁵ As a next step, the Treasury and the Department for Business, Innovation and Skills presented a White Paper entitled *Banking Reform: Delivering Stability and*

49 ICB, 'Issue Paper: Call for Evidence' (ICB 2010) 34.

50 ICB, *Final Report: Recommendations* (ICB, September 2011).

51 Ibid 48.

52 *Financial Times* (London, 29 August 2011).

53 G Mathewson 'Ringfencing not Real Threat to Britain', *Financial Times* (London, 11 September 2011).

54 Financial Services (Banking Reform) Bill, Bill 002 2013–14.

55 <www.moneymarketing.co.uk/regulation/-rbs-chief-says-customers-will-pay-ringfencing-costs/1051312> accessed 14 May 2012.

Supporting a Sustainable Economy that opened the next round of public consultation. Besides setting out the context of the reforms, the White Paper discusses ring-fencing, loss-absorbency and competition in separate chapters. It also presents in the annex an impact assessment comparing the benefits and costs of the Coalition government's implementation plan of the IBC recommendations with a reference policy option of non-action.

Vickers criticised the government saying that the proposal would not go far enough and fall short of fully implementing the recommendations of the IBC, especially watering down the limits on the leverage of big banks.⁵⁶

Throughout the debates initiated by the establishment of an IBC and the consultation processes on an initial call for evidence and then on the interim report, there were warnings that the UK should not adopt any approach unilaterally but 'should lead a global debate on structural reform'.⁵⁷ Industry was generally in opposition of any limits on proprietary trading and investing, arguing for market efficiency and the need for hedging and market-making.⁵⁸ Barclays⁵⁹ argued that there would not be any evidence suggesting 'that structural change would enhance financial stability', the costs for structural change would be too high and implementation uncertain. Structural changes with regard to diversity of business models, size and types of banks should be left to the market and the institutions themselves. Any changes should conform to international reforms with regard to scope and timeline. The British Bankers Association (BBA)⁶⁰ defended universal banks, emphasising their key role in economic recovery. Moreover, the BBA warned 'that separately devised national measures or international measures introduced on a super-equivalent basis in the UK will not only have the potential to be detrimental to the UK's attractiveness as a financial centre, but also be damaging to the economic future of the UK'. HSBC⁶¹ advocated the universal banking model as 'best placed to meet 21st century needs'. Moreover they would be 'more stable' than other forms of banks. In summary these arguments are the pre-crisis arguments claiming advantages of financialisation.

Following the interim report and the final report, industry remained concerned about the ICB's ring-fencing proposal. Critique included the argument that it would be too costly with a negative impact on recovery, that it would increase the stress on the unprotected entity in times of crisis,⁶² or that ring-fencing should not be designed in a way that it has impacts on any of the current 'standard banking services demanded by their individual and business customers of any size'.⁶³ The BBA has also been concerned about ring-fencing and demanded much more evidence and research on 'the effect of such a proposal on financial stability under different economic scenarios' as well as 'the consequences for the internationality of the UK as a financial centre should such a proposal proceed'.⁶⁴

This is all evidence of a weakened banking industry. However, Angela Knight, chief executive of the BBA, eventually welcomed the banking reform White Paper stating that it

56 *Guardian* (London, 14 June 2012).

57 Augar, Evidence to the ICB, 15 November 2010. All following evidence papers are available on the ICB website <<http://bankingcommission.independent.gov.uk/responses/>>.

58 Eg Association for Financial Markets in Europe evidence to the ICB, 19 November 2010.

59 Evidence to the ICB, 15 November 2010.

60 BBA, evidence to the ICB, 15 November 2010.

61 Evidence to the ICB, 24 November 2010.

62 See eg the Royal Bank of Scotland's response to the Interim Report, 4 July 2011. The responses to the ICB Interim Report are available at the ICB website <<http://bankingcommission.independent.gov.uk/interimresponses/>>.

63 Santander Response to the Interim Report (n 62).

64 BBA Response to the Interim Report (n 62).

'balances the need to establish a banking system which is more certain and more secure with the ability of banks to continue to provide the full range of services to customers and support to the economy'. The BBA especially welcomed 'the refinements to the original proposal', which in total would 'have removed many of the obstacles from the original [ICB] report'. Although 'the costs to the banks' would still be 'substantial', Knight no longer saw a risk for the provision of services or the City's international position.⁶⁵

Since then the Parliamentary Commission on Banking Standards has concluded that the ring-fence proposals are insufficient, because there would not be sufficient precautions to ensure that the ring-fence does not get holes over time via lobbying and political interventions. The Parliamentary Commission therefore suggested 'electrification' of the ring-fence. The regulator would need strong powers to enforce separation of institutions if they do not meet strict ring-fence requirements. This would also require periodic reviews of the sector, and over time even tougher measures might be required. As next task the Parliamentary Commission wanted to look into full separation of proprietary trading along the Volcker Rule line.⁶⁶ These proposals go in an interesting direction from the perspective of gradual institutional change and address typical problems of financial reforms becoming less strict in implementation and enforcement over time.

A slightly different approach to ring-fencing comes from the EU's High-level Expert Group on reforming the structure of the EU banking sector. The group was set up by Commissioner Michel Barnier in November 2011 to assess the need for bank structural reforms and it was chaired by Erkki Liikanen, Governor of the Bank of Finland. The group held a consultation in May 2012 to which mostly banks and bank associations but also Finance Watch, Friends of the Earth Europe, and Transparency International responded. Contributions from megabanks were largely missing. The group's final report, released on 2 October 2012, proposed a mandatory separation of certain trading-related activities (above a specific threshold) which would require banking groups to organise certain businesses into a separate legal entity that must be separately capitalised and must not be funded by insured deposits.⁶⁷ In contrast to the ICB position, the idea here is to ring-fence the investment banking arm with the trading and market-making arms.

The Liikanen Report was welcomed by Commissioner Barnier and was put out for a six-week consultation. Among first reactions, industry representatives welcomed that the report did not put into question the universal banking model. But the hybrid between Volcker and Vickers was also seen as problematic by many industry representatives. Some argued that this reform could damage the flow of corporate funding in Europe.⁶⁸ The next months will show reactions to the report and what will be translated into new regulations. In the meantime, this approach has been followed by Germany. The German Banking Separation Act, to be implemented by mid-2015, has, however, similar exemptions to the Dodd-Frank Act with regard to market-making and dealing on behalf of customers. The German approach might guide the direction for other countries with a strong universal banking tradition. In summary, the developments on ring-fencing led to policies to modify the existing structures only moderately and to keep important loopholes that do not undermine

65 <www.bba.org.uk/media/article/angela-knight-banking-reform-white-paper-balance-safety-with-the-needs-of-g>, 14 June 2012.

66 House of Lords—House of Commons Parliamentary Commission on Banking Standards, *First Report Together with Formal Minutes*, vol I (HL 2012 98 HC 2012 848) (The Stationery Office 2012).

67 E Liikanen et al, *High-level Expert Group on Reforming the Structure of the EU Banking Sector: Final Report* (Brussels, 2 October 2012).

68 *Financial Times* (London, 8 October 2012).

the dominance of financialisation. Loopholes and flaws are similar to the US approach and can be explained by massive business pressure and fears about competitive disadvantages.

3.4 NARROW BANKING

A third proposal to end bailouts of financial institutions is narrow banking. This idea has its origins in the US where it was previously considered in the 1930s. It is based on the belief that the fractional reserve banking system would have significant problems and weaknesses and that it would only work in normal times. Whenever faced with extraordinary demand for cash, banks would necessarily face bank runs, break down or require bailouts. In the 1930s, economists at the University of Chicago proposed the ‘Chicago Plan’ and Yale economist Irving Fisher suggested a ‘100 percent money’ system.⁶⁹ The core idea of these proposals was to split the banking system into a protected commercial banking system and a risky investment banking system. Milton Friedman also became an advocate of narrow banking when he updated the Chicago Plan.⁷⁰ Since then, narrow banking has remained a topic that again and again gained some attention. In the 1980s, Robert Litan of the Brookings Institution proposed narrow banking to deal with moral hazard.⁷¹ His idea was to create monetary service companies that would only hold safe assets. Ronnie Phillips discussed not just the advantages of narrow banking but also engaged with critics.⁷² Another important study was an International Monetary Fund working paper entitled ‘Should Banks Be Narrowed?’ that argued that narrow banking would result in ‘less than certain benefits in terms of greater financial stability, while it would exact some heavy costs in terms of efficiency and credit availability’.⁷³

Now, in the aftermath of the financial meltdown, narrow banking has again found some wider support. Phillips and Roselli proposed narrow banking for the US.⁷⁴ John Kay, a Scottish economist, professor at the London School of Economics and *Financial Times* columnist, is the strongest supporter of narrow banking in the UK. According to Kay, the purpose of narrow banking ‘is to protect the non-financial sector as far as possible from the consequences of failures and instability within the financial sector’. This would only be possible by a complete separation of ‘the casino from the utility’.⁷⁵

Advocates of narrow banking share the belief that better or more regulation will not prevent financial institutions from excessive risk-taking. Supervision would be the wrong approach and narrow banking would require only limited regulation. Kay, for example, stated that the principles should be ‘competition where possible, regulation where necessary, and supervision not at all’.⁷⁶ This distaste of supervision comes from the belief that it is necessarily resulting in regulatory capture.⁷⁷

69 I Fisher, *100% Money* (City Printing Company 1935); I Fisher, ‘100% Money and the Public Debt’ [1936] (April–June) *Economic Forum*.

70 M Friedman, *A Program for Monetary Stability* (Fordham UP 1960).

71 R E Litan, *What Should Banks Do?* (The Brookings Institution 1987).

72 R Phillips, *Narrow Banking Reconsidered: The Functional Approach to Financial Reform* (Levy Institute Public Policy Brief, no 18/1995) <www.levyinstitute.org/pubs/ppb17.pdf>.

73 B Bassone, ‘Should Banks Be Narrowed?’ (IMF Working Paper WP/01/159 2001).

74 R Phillips and A Roselli, *How to Avoid the Next Taxpayer Bailout of the Financial System: The Narrow Banking Proposal* (Networks Financial Institute Policy Brief 2009-PB-05) available at SSRN <<http://ssrn.com/abstract=1459065>>.

75 J Kay, ‘Narrow Banking: The Reform of Banking Regulation’ <www.johnkay.com/wp-content/uploads/2009/12/JK-Narrow-Banking.pdf>; J Kay, ‘Should We Have “Narrow Banking?”’ in Turner et al (n 19).

76 Kay, ‘Should We Have “Narrow Banking?”’ (n 75) 212.

77 Ibid 210.

The idea of narrow banking found little support in regulatory reforms. It was nevertheless discussed in financial media, several economists expressed sympathies and also the UK ICB looked into it and asked for comments from stakeholders. In its 2010 'issue paper' the ICB framed the debate with several concerns against narrow banking, such as lost synergy effects between deposit-taking and bank lending activities, a 'tying up' of 'banking deposits in government bonds' instead of making them 'available to fund productive investment through financial intermediation'. Moreover, narrow banking would not end the possibility and need for bailouts.⁷⁸ In the final report, the ICB argued that narrow banking 'would dramatically reduce the range of intermediation possibilities' and that 'ring-fencing would be less costly'.⁷⁹ Kay himself discusses and rejects a number of critiques such as the claim that the world would need big banks, that big banks would be superior because of their ability to reduce risks by diversification, or that corporate customers would demand universal banking. He also opposed the claim that the unilateral introduction of narrow banking would result in competitive disadvantages. He also argues that 'living wills' and a proper resolution regime for insolvent banks would require 'radical restructuring and simplification of the corporate structures of financial conglomerates', resulting in narrow banking.⁸⁰

However, throughout the ICB consultations there was no big support for narrow banking. HSBC dealt with narrow banking in an annex and claimed that it 'would be a retrograde step' and that it would be 'a primitive and costly model that would cause the cost of lending to rise significantly'. The concept would also ignore the market demand for 'more sophisticated products'. As a result, the corresponding risk of such products would simply shift 'somewhere else in the system'.⁸¹ In short, narrow banking is an idea that would lead to a radical transformation of banking but which would require a radicalisation in the overall political debates for realisation. Then it would still compete with other radical ideas.

3.5 LIMITED PURPOSE BANKING

Another radical minority demand to end casino banking is the introduction of limited purpose banking. Here the argument is that narrow banking would not go far enough and would still invite 'all the problems of shadow banking' but that there would be a very simple way of making the banking system safe and ending financial collapse and contagion by transforming the system in line with the experience with mutual funds. Laurence Kotlikoff, professor of economics at Boston University, advocates the idea.⁸² He argued in an open letter to Sir John Vickers, published in the *Financial Times* that:

Under Limited Purpose Banking, there are no incorporated shadow banks because all incorporated financial companies, whether they call themselves commercial banks, investment banks, hedge funds, SIVs, insurance companies, etc., are forced to operate in one way only – as mutual fund companies, which issue mutual funds with absolutely zero leverage, including no short positions.⁸³

Financial companies would need to decide whether they operate with limited or unlimited liability. Under limited liability, firms would operate 'as 100 per cent equity-financed mutual funds' and they would only be allowed to engage in activities that do not involve leverage.

⁷⁸ ICB (n 49) 33ff.

⁷⁹ ICB (n 50) 137.

⁸⁰ <www.johnkay.com/wp-content/uploads/2009/09/Narrow-Banking-FAQs-Jan2010.pdf>.

⁸¹ Evidence to the ICB, 24 November 2010 (n 57).

⁸² L J Kotlikoff, *Jimmy Stewart Is Dead* (John Wiley & Sons 2010); L J Kotlikoff and E Leamer, 'A Banking System We Can Trust' *Forbes*, 23 April 2009 <www.forbes.com/2009/04/22/loan-mortgage-mutual-fund-wall-street-opinions-contributors-bank.html>.

⁸³ *Financial Times* (London, 28 January 2011).

As their purpose is limited to financial intermediation, risk-taking shifts from the companies to individuals.

Kotlikoff assumes significant consequences of such a change, including the disappearance of shadow banking and risky speculation of firms with unlimited liability when ‘the ‘brilliant’ titans of Wall Street who produced the financial debacle’ would have to fear losing ‘their homes, villas, yachts, Austin Martins, etc’.⁸⁴ He also claims that a mutual funds system could never fail and that such institutions would only act, as intermediaries as kind of ‘honest middle men’. In his view, the problem of modern finance is not about complex financial products but about systemic large-scale, multifaceted fraud and lack of information disclosure. These problems could be overcome by limited purpose banking and a federal regulator assuring strict control and full, real-time information disclosure. This new federal regulator he labels Federal Financial Authority (FFA) and is designed similar to the US Food and Drug Administration. The FFA would rate financial securities, the safety of investment banks, insurance companies, hedge funds and commercial banks.⁸⁵

Limited purpose banking was also a topic in the UK ICB consultations. The 2010 issue paper stated that this alternative ‘is open to some of the same criticism as narrow banking’, that it would be too costly and that participating households would be negatively affected by the increase of illiquid assets. HSBC argued that the already existing limited purpose banking would show that it ‘depends on the existence of a modern banking system’: ‘Without banks to originate the loans, package them into securities and distribute them to mutual funds, “limited purpose banking” could not take place’. Moreover, HSBC recalled weaknesses of money market mutual funds that had come to light over recent years and been addressed by the Dodd-Frank Act.⁸⁶ In short, this is another idea that would require a completely different situation to gain more attention and will definitely not come via any form of gradual change.

3.6 NATIONALISATION OF BANKS AND DEMOCRATIC CONTROL

Over the past few years quite a number of troubled banks have had to be nationalised. Politicians were more than keen to emphasise that this was an extraordinary and temporary measure that should not be perceived as going back to pre-1970s state industries approaches. Nationalisation was the emergency step in a situation in which no private investor was willing to take over the risk. Alan Greenspan, for example, argued ‘that once in a hundred years this is what you do’.⁸⁷ State actors underlined that the nationalised banks should be privatised as soon as the possible. Since then a debate has started over whether the state should use this new situation to gain a maximum return by keeping them much longer as state property or by using them as examples for transforming banking in a broader societal public interest direction. Additionally to nationalisations, massive debts of the private sector have been socialised and transformed into public debt. Some commentators, such as Martin Wolf, have argued that all the bailouts have already abolished the state–private divide and made banking strongly linked to the state which would legitimise the state to intervene much more aggressively. The most radical proposal, however, is the nationalisation – or better socialisation – of banks and putting them under democratic control. This is a demand, by some in the Occupy and labour movements and by Marxists, that signifies a major transition of existing varieties of capitalism and is normally linked with a less evolutionary and more revolutionary socialist or communist transitional

84 *Financial Times* (n 83).

85 <www.kotlikoff.net/content/interview-limited-purpose-banking>.

86 Evidence to the ICB, 24 November 2010 (n 57).

87 *Financial Times* (London, 18 February 2009).

programme. But environmentalists have also been arguing for societal control over banking to incorporate sustainability principles. Historically, Marx and Engels called for ‘centralisation of money and credit in the hands of the state, by means of a national bank with state capital and an exclusive monopoly’, meaning ‘the suppression of all private banks and bankers’. This measure was supposed ‘to regulate the credit system in the interest of the people *as a whole*’ and to ‘undermine the dominion of the big financial magnates’.⁸⁸ The demand was then a key aspect in the Bolshevik programme formulated in Lenin’s April Theses and in his brochure on ‘the impending catastrophe and how to combat it’, both from 1917. The April Theses called for the ‘immediate union of all banks in the country into a single national bank, and the institution of control over it by the Soviet of Workers’ Deputies’.⁸⁹ In the ‘impending catastrophe’, Lenin argued that the nationalisation of banks and their concentration into a single state bank is the only measure allowing a democratic control over finance given its complexity and the various accounting tricks. It would allow for the possibility ‘to *regulate* economic life’ and allow economic development without ‘sky-high’ profits for bankers. ‘Banks nowadays’, he stated, ‘are so closely and intimately bound up with trade (in grain and everything else) and with industry that without “laying hands” on the banks nothing of any value, nothing “revolutionary-democratic”, can be accomplished’.⁹⁰

The nationalisation of banks and their transformation into a single state bank was seen as an easy task to be achieved by a simple decree and ‘carried out by the directors and employees themselves’. Bankers resisting this change ‘would lose their highly remunerative posts and the opportunity of performing highly profitable fraudulent operations’. It would also be necessary to confiscate their property and to imprison bankers, board members and big shareholders if they tried ‘to conceal documents and accounts’. According to this transition programme, there would still be a market and the new banking structure would provide ‘credit on easy terms for the small owners’ and ‘for the peasants’.⁹¹

It is obvious why such a programme should become to a certain extent popular again in a situation of deep economic crisis with high unemployment and unpopularity of bankers and why nationalisation and democratic control could at least theoretically be seen as necessary for a policy in the interests of the majority. Of course, proponents have drawn lessons from banking in eastern European countries under Soviet influence. Trotskyists still argue in line with Lenin and the debates of the early 1920s for a state monopoly under democratic workers’ control. Other left-wing groups no longer call for a state monopoly and some have even wondered if the property question is relevant at all in a modern context and if democratic control could actually be achieved over private banks.

Nationalisation and democratic control have been demanded by a number of organisations, such as very small left-wing groups, but also more influential ones such as Attac, the Occupy movement, or Die Linke (The Left) in Germany. Sarah Wagenknecht, representing the Marxist wing within The Left, has argued for these demands over the last few years since she got involved in the regulatory reform debates within the European Parliament and gained some wider public attention in the German discourse. The party programme, adopted in 2011, very much follows her earlier line by calling for ‘a banking

88 K Marx and F Engels, *Manifesto of the Communist Party* (1848) <www.marxists.org/archive/marx/works/1848/communist-manifesto/index.htm>.

89 V I Lenin, ‘The Tasks of the Proletariat in the Present Revolution [aka The April Theses]’ (1917 <www.marxists.org/archive/lenin/works/1917/apr/04.htm>.

90 V I Lenin, ‘The Impending Catastrophe and How to Combat It’ (1917) <www.marxists.org/archive/lenin/works/1917/ichtci/index.htm>.

91 Ibid.

system with three pillars: savings banks, cooperative banks and large state banks'. 'A functioning financial sector', the programme continues, 'is a public good, so providing it is a public task.' Banks should be mandated to follow 'social and ecological guidelines' and 'to serve the public weal'. In their governance structures they should moreover incorporate 'guaranteed strong codetermination rights as a corrective to management decisions'. Direct citizens' involvement is proposed with regard to 'the development of municipal services'. Public property of the large banks would assure a 'new economic management' and political priority setting. 'Private banks must . . . be nationalized, subjected to democratic control and made to serve the public weal.' The 'public service mission' must be protected by 'strict regulation'. Investment banking must 'be phased out, it must be forbidden for banks to trade in securities for their own account and to speculate in derivatives, as well as to transact any business outside of their own balance and to have any dealings with companies or persons legally registered in tax havens'. Hedge funds and private equity firms should be banned, capital controls reintroduced and various speculative products banned.⁹²

All these proposals are not elaborated in any detail with regard to how 'democratic control' or 'workers' control' should work. However, they raise a question that addresses how practices and activities of finance could be brought into line with a wider public interest. They also take into account that financial stability and many activities banks should provide are public and no longer private. It goes without saying that these proposals have not been taken up by any regulators and that they are marginalised views. But, even from a less radical perspective, there are a number of lessons to be drawn from the nationalisations of the crisis years so far, especially with regard to repeated bailouts. Regulators need to be better prepared for nationalisations and must take much stricter requirements into account before they nationalise a bank. This is a debate that is still largely underdeveloped in the current turmoil and preoccupation with crisis management. A complete socialisation of banks and democratic control, however, would definitely require a revolution or an enormous shock such as a break-up of the Eurozone or another massive aftershock; this could not possibly come into existence via gradual institutional reforms.

3.7 RECOVERY AND RESOLUTION PLANS

More prominently on the agenda are demands for the orderly resolution of failing institutions. The regulatory reforms adopted or discussed so far have focused on unwinding 'SIFIs and 'globally systemically important banks' (G-SIBs) via so-called living wills, funeral plans and orderly resolution systems. This policy is a reaction to the lesson that the current system with all the interconnectedness and complexity makes it nearly impossible to let failing banks go bankrupt. In short, it tries to end the TBTF and 'too-interconnected-to-fail' problem and the survival of highly undercapitalised 'zombie banks'. The core ideas were agreed on at G20 level, the Financial Stability Board provided guidance, and the policy has now to be implemented. The US adopted this approach in the Dodd-Frank Act (Title I, s 165(d)) and has already started implementation. A first group of megabanks had to submit their funeral plans by mid-2012, with a second group following by mid-2013.

The EU is lagging behind and plans have been delayed due to the crisis at member-state level as well as in Brussels. The European Commission presented a proposal in June 2012⁹³ that is now being debated.

92 Die Linke, *Programme of the Die Linke Party* (Die Linke 2011).

93 A Bank Recovery and Resolution Directive (COM(2012) 280/3) together with a first impact assessment (SWD(2012) 166 final).

There are several serious problems with these approaches. Firstly, they require frequently updated public and confidential reports with the latter only being seen by the regulatory authorities. The overall soundness and effectiveness rests with the quality of the confidential reports and the willingness of regulatory authorities to enforce a strict policy. Given the historic experiences with regulatory capture and foreseeable problems of strict implementation in good times, the current designs are not very promising. Insiders from the US have argued that the initial experience is rather disappointing and that the banks' living wills were not of a high standard. This might change via gradual institutional reforms and learning processes over time. Secondly, there are serious doubts about how living wills would actually work in a crisis and whether this might increase capital concentration.⁹⁴ Thirdly, there are doubts if the idea works at all as long as all the banks are so interconnected.⁹⁵ Yet there is also the problem of US regulators being able to shut down smaller institutions but not the larger megabanks because Dodd-Frank only applies in the US and does not cover cross-border operations.⁹⁶ In short, there are still many problems to be solved with regard to the orderly resolution of failing banks.

4 Conclusion

This paper discussed why structural reforms of finance became so prominent with the global financial crisis and why structures need to be changed. The adopted regulatory reforms with a direct effect on financial structures that have been analysed have introduced significant changes, which, however, will not transform the current system of financialisation. They have, of course, to be seen in the context of the wider reforms that are taking place, especially indirect regulation affecting the structures of finance or product regulations that were beyond the focus of this paper. What can be learned from direct regulation is that overly complex rules were created with the purpose of generating flaws and loopholes that do not undermine the dominance of finance created by overly complex and interconnected sectoral structures. Foremost, these rules made the divergence between the US, with the adoption of the Volcker Rule, and the UK and the EU, with their proposals for ring-fencing systems, less significant by ensuring no threat to competitive advantages in the corresponding domestic systems. It is still too early to anticipate the overall effects the reforms will generate on regulatory competition in the long run and in particular on the financial centres' futures. What is already clear, however, is that the complexity of the regulatory reforms allows for a number of problematic aspects and trends to continue. The policy window for a radical system overhaul has not been used so far. Complexity and uncertainties as well as highly technical legal debates and existing regulatory governance practices give moderate reform proposals an advantage and leave radical left proposals marginalised. The contemporary practice of the better law-making approach with stakeholder consultations and impact assessments made it easy to adopt only proposals reflecting dominant powers and interests. The framing of such processes as well as the difficulties in calculating long-term effects of reform proposals and the lack of corresponding studies were identified as essential in several debates. A radical overhaul of financialisation is difficult to achieve without more public pressure and stronger countervailing powers. Aftershocks or the next level of the crisis could easily bring more radical proposals onto the agenda. These more radical regulatory ideas are not all new but

94 J Silver-Greenberg and N Schwartz, "Living Wills" for Too-Big-to-Fail Banks Are Released', *New York Times* (New York, 4 July 2012).

95 J Eisinger, 'Banking Regulator Calls for End of "Too Big To Fail"', *New York Times* (New York, 28 March 2012).

96 S Johnson, 'The Myth of Resolution Authority', *New York Times Economic* (New York, 31 March 2011).

build on longer traditions as shown. They have been adopted and modified to new situations over time. Nevertheless, they are mostly still underdeveloped and lack clear transformative strategies that could make them reality. Some ideas like narrow banking or limited purpose banking are more of a minor expert discourse, while nationalisation seems to be a more natural response driven by different motivations in specific situations that could quickly spread. Most influential on actual reforms were, however, warnings against over-regulation and ideas about structures mainly created by market developments. Given the strong pressures from industry in the overall climate of crisis management and recovery, it is a success that the adopted structural reforms are at least making the overall system more preventive compared to the pre-crisis system. Whether gradual institutional reforms will improve institutional arrangements and whether strict implementation and enforcement can be ensured over time is still uncertain and will need close attention by regulators and researchers.

