NORTH EAST LAW REVIEW

NEWCASTLE ★ EST. 2013 ★ UNIVERSITY

Volume 1
2013

North East Law Review

Volume 1

June 2013

North East Law Review

Newcastle University

Board of Editors

Principal Officers

General Editor	Nikita Beresford
Managing Editor	Catherine Caine
Treasurer	Jessica Randell

Constitution Committee

Assistant General Editor	Foteini Kyriakopoulou-Kollia
	Thomas Gordon
	Emily Owen

Style Guide Committee

Assistant General Editor	Mary Wong
	Ian Mckie
	Anamaria-Mitina Mihaita

Sponsorship Committee

Alice Deeley
Ryan Hodgins
Hollie Morgan

Website Committee

Website Manager	Lida Pitsillidou
	Emmanuel Tayari
	Aleksandra Wawrzyszczuk

Foreword

Message from the General Editor

I would like to welcome you to the inaugural volume of the North East Law Review. What follows is the product of the sustained and admirable hard work of contributors and editors, and I would like to take this opportunity to say thank you to each and every person who helped in the creation of this review. Firstly, to the article writers for generously allowing us to use their thoughts, ideas and words; the quality is undoubtedly high and reflects wonderfully on the studentship of Newcastle University. Secondly, the Editorial Team, who have worked tirelessly not only in preparing the journal itself, but also in stepping boldly into the unknown and forming the North East Law Review, bringing it from nothingness to a successful and sustainable committee in just a few short months. I am overwhelmed by their achievements. Thirdly, to our Staff Liaison, Colin Murray, who provided the initial ideas, and gave constant encouragement and guidance – without him, success would not have been possible. I would also like to thank the Head of School, Chris Rodgers, for his support, and the administrative and library staff for their assistance. Finally, I would offer my greatest thanks to the reader of this edition – I hope we live up to your expectations, and that you will pick up a copy of the NELR again.

Nikita Beresford

Message from the Managing Editor

It has been a great pleasure to be acting as the Managing Editor of the North East Law Review in its first year of establishment. I believe that this first volume is a great testament to the hard work of the student contributors from Newcastle University and the student Editorial Board. The North East Law Review was established with the aim of providing an opportunity for student writers within the UK to present their work to, and engage in debate with, the academic community, and in this respect, I believe that the NELR is unique. Providing students with an accessible platform to publish their work is something that I am passionate about from the challenges faced in trying to get my work in environmental law published. I believe that the North East Law Review can provide an inspiration to students from other Universities to create their own Law Reviews, and encourage law students to write with a view to getting their work published.

Catherine Caine

Table of Contents

‘I AM RICHARD II, KNOW YE NOT THAT?’ – WORKS OF WILLIAM SHAKESPEARE AND THE ART OF GOVERNANCE

HENRY JONES[*]

William Shakespeare had a great impact upon the early modern English society. His works touch upon the art of governance, responsibilities of princes and the importance of their conduct as subjects of law, regardless their factual public role. This article contrasts two characters appearing in Shakespeare’s plays: Princes Hal and Richard. It demystifies Shakespeare’s opinions on good conduct, and presents the responsibilities of monarchs from a theatrical context. The article considers how Shakespeare’s theatrical productions reflected both anxieties of late 16th century England, and Shakespeare’s contemporary attitudes to an ideal model of governance.

This article considers Shakespeare’s commentary on the responsibility of princes and the nature of governance in the context of early modern England. It will find that Shakespeare’s work reflects and crystallises the contemporary attitudes of late 16th century England, providing an insight into how royals were supposed to act and govern. This will illustrate that princes must become well versed in the art of governance when presenting the role of the public figure. The article will first discuss the historical influences and attitudes which provide the context to Shakespeare’s commentary. Secondly, it will examine the education and responsibility of princes by focusing on prince Hal’s development in *King Henry IV Part 1*. Finally, it will consider the nature of governance in relation to constitutional law by analysing the role of Richard in *King Richard II*. The article will find that a ruler must learn to conduct themselves as subjects of the law, even if their public role elevates them above this status.

Shakespeare wrote during a time of dramatic vicissitude and instability throughout society. One aggravating factor was that, partly due to a ‘series of poor harvests ... the 1590’s was a decade of general economic difficulty’.[1] Religious stability was also doubtful; ‘only a half-century had passed since Henry VIII first separated the Church of England from Rome ... it was still most uncertain what kind of authority it could have’.[2] Elizabeth I had used the new religion as an instrument of power. Protestantism had endowed the population with the belief that England was a nation supported by God. The notion of the ‘godly commonwealth was prescribed by late Elizabethan

[*] Henry Jones, Newcastle University, Law LLB Stage Three.

[1] Ian Ward, *Law and Literature: Possibilities and Perspectives* (first published 1995, Cambridge University Press 2008) 61.

[2] James Boyd White, *Acts of Hope: Creating Authority in Literature, Law, and Politics* (University of Chicago Press 1994) 82.

political theory'.[3] This attitude was affirmed by national successes such as the 1588 English defeat of the Catholic Spanish Armada, and works such as Hooker's *Of the Lawes of Ecclesiastical Polity*. However, towards the end of Elizabeth's reign, society's uncertainties were exacerbated by the fear of change after her succession.

Shakespeare's works capture the concerns of late 16th century England; he tackles contemporary issues, such as the waning of Elizabethan authority. It is fitting that his social commentary is dramatised because authority itself was, and still is, a performance. As a leader 'Elizabeth always sought to "dazzle" .... Her incessant royal progresses around the commonwealth were designed to put her and her court on show'.[4] The monarch had a hugely important public duty so naturally their public persona played a significant role. However, this could be misplayed. For instance, it may be that society's concerns about Elizabeth's succession were justified as 'James was a strikingly inept performer ... writing explicit love-letters to young male courtiers'[5] was one of his many downfalls. Shakespeare's characters such as prince Hal and Richard II address these public successes and failures.

Shakespeare's depictions of these characters had a socio-political 'agenda, as well as an audience in mind'.[6] Ward notes that in this kind of text, 'to ignore the author ... is to be deliberately obtuse'.[7] It is therefore acceptable to consider Shakespeare's work as a commentary which presents concerns and problems of the time. As such we can see characters like Hal and Richard II, as being examples of good and bad use of authority for a 'godly commonwealth' to draw upon. In this sense, Shakespeare offers the reader 'ways of imagining the world and claiming authority within it'.[8] His work encouraged understanding of its themes. However, this commentary is not explicit, rather it is presented as part of the performance, 'powerfully conveying how eloquence can either serve or distort justice'.[9] As Weisberg suggests, the presentation of a legal problem 'places the reader in the position of a juror, who is thus engaged creatively with the text'.[10] While Shakespeare's text can be drawn upon by the reader or audience, one must consider that 'anything, once a sufficiently elaborated argument is in place, can mean anything'.[11] Therefore, it is important to remember that Shakespeare's characters, while guiding a social commentary, are not purely instructional and should be considered as part of the production.

The example of the development of a responsible princely conduct is evident with the education of prince Hal in *Henry IV Part 1*; a role which gradually answers the question:

---

[3] Ian Ward, *Shakespeare and the Legal Imagination* (Butterworths 1999) 51.
[4] ibid 200.
[5] ibid 200.
[6] Ward (n 1) 35.
[7] ibid 36.
[8] Boyd White (n 2) 50.
[9] Richard Weisberg, *Poethics* (Columbia University Press 1992) 40.
[10] Ward (n 1) 38.
[11] Stanley Fish, *There's no Such Thing as Free Speech: and It's a Good Thing too* (Oxford University Press 1994) 146.

'what are the ... qualities that go to the making of a king?'[12] We first encounter Hal in conversation with Falstaff, a drunkard and criminal resident of Eastcheap. While Hal associates with Falstaff's people, he is clearly not one of them, and is often quick to distance himself from Falstaff with sharp insults; 'Thou art so fat-witted, with drinking of old sack'.[13] This is young Hal before he has learnt the public art of governance, yet he is clearly princely in his private opinions. For instance, when a robbery is suggested by Poins - another criminal, Hal initially responds; 'Who-?-I rob? I a thief? not I, by my faith'.[14] Later, when Hal is alone, he explains that he only associates with Falstaff's people in order to better impress his critics when he is educated; 'My reformation, glitt'ring o'er my fault, Shall show more goodly and attract more eyes'.[15] This alludes to the fact that he understands the duty to perform as a prince to the public. Yet Hal does not have the direction or understanding to become a leader, he still has important lessons to learn.

As Hal learns, the audience in turn is informed by Shakespeare's commentary upon how a prince should act. In act two, scene four, his behaviour is immature and ugly. During a role-play between Hal and Falstaff, Hal plays his father Henry IV and deplores Falstaff as 'that bolting-hutch of beastliness ... that huge bombard of sack'.[16] This is a cruel deconstruction of 'Shakespeare's most popular character'[17] that does not endear Hal to the audience. This scene also sees Hal acting above the law; he lies to the Sherriff that Falstaff, the 'gross fat man',[18] is not in the tavern before pick-pocketing him while he sleeps. While Traversi says that this condemnation is a 'look forward to the final rejection' of Falstaff,[19] Henry IV is still furious with Hal for fraternising with the scoundrels of Eastcheap. Hal is disciplined into acting lawfully and dutifully by Henry in act three, scene two.

Here Henry compares Hal to the tyrannical Richard II. He describes a man who indulged in bad company; 'the skipping King, he ambled up and down with shallow jesters and rash bavin wits, soon kindled and soon burnt'.[20] A warning that Hal should take on a more responsible role, rather than follow 'Richard's failings, here so forcibly and truly described, amount to a degradation of royalty'.[21] This underlines that the company an aspiring prince keeps, and the advice he takes is essential to effective leadership. In Hal's case, 'the most immediately destabilising constituent of this aspiration seems to be Henry's affinity with the "good lads" of Eastcheap'.[22]

---

[12] Derek Traversi, *Shakespeare: From Richard II to Henry V* (first published 1958, Hollis & Carter 1979) 4.
[13] William Shakespeare, 'Henry IV Part 1' in Peter Alexander (ed), *William Shakespeare: The Complete Works* (first published 1951, Collins 1985) 1.2.3-4.
[14] ibid 1.2.133-134.
[15] ibid 1.2.206-207.
[16] ibid 2.4.436-437.
[17] Michael Freeman (ed), *Law and Popular Culture* (Oxford University Press 2005) 5.
[18] Shakespeare (n 13) 2.4.492.
[19] Traversi (n 12) 76.
[20] Shakespeare (n 13) 3.2.60-62.
[21] Traversi (n 12) 83.
[22] Ward (n 3) 55.

Shakespeare subscribed to the Sir John Fortescue LCJ's views on constitutional law that 'governance was shared by the King and a very few select subjects'.[23] He illustrates to the audience the importance of a prince keeping good company and remaining subject to the law.

Hal does learn from Henry, and his princely education flourishes. In act three, scene three he makes amends for his crude behaviour at the tavern. He informs Falstaff of his pick pocketing, and makes sure 'the money is paid back again'[24] after the robbery of the day before. This indicates Hal's new approach, as a good prince and as a subject of the law. From here Hal goes from strength to strength; in act five, scene one, we see his skill in diplomacy as he offers a potential solution to 'save the blood on either side'[25] in battle. Although this plan is foiled, Hal goes on to fight bravely. The audience is reminded of the importance of his education when Hal encounters the cowardly Falstaff in battle; 'What, is it a time to jest and dally now?'[26] The prince has a maturity and responsibility, important to Shakespeare and his audience. Prince Hal has set the example for England to follow.

This exemplary behaviour is taken further in *Henry V* as Hal becomes a strong king of England and an ideal role model for its people. In *Henry V*, Shakespeare's commentary is clear through the voice of the chorus who are 'dedicated to presenting the imaginary, and performative, nature of an ideal commonwealth'.[27] By presenting the example of Hal's development in this way, it is impossible to ignore Shakespeare's voice and commentary upon princely education. He uses bad influences such as Falstaff, and good influences such as Henry IV to convey this example to the audience, and allow them to reach a conclusion. In this regard, one could apply a Bakhtinian understanding that Shakespeare's work is polyphonic in that he 'incorporates many different styles, or voices, which as it were talk to each other, and to other voices outside the text, the discourses of culture and society at large'.[28] Although Bakhtin uses the idea of polyphony in relation to novels, Lodge suggests that 'it would not be difficult to construct a Bakhtinian reading of Shakespearean drama, which is manifestly polyphonic'.[29] Shakespeare presents another voice of society in *Henry IV Part 1*, that of the chaotic world of Eastcheap and its inhabitants.

It is debatable whether or not the world of Eastcheap is part of a wider political discussion. It is presented as 'a world of imagination, literally run riot. The commonwealth of England seems to be almost entirely fictitious.'[30] It could be that 'Shakespeare, as Melchiori suggests, was first and foremost a man of the theatre, not of

[23] Ward (n 1) 63.
[24] Shakespeare (n 13) 3.2.177.
[25] ibid 5.1.99.
[26] ibid 5.3.52.
[27] Ward (n 3) 62.
[28] David Lodge, *The Art of Fiction: Illustrated from Classic and Modern Texts* (Penguin 1992) 128.
[29] David Lodge, *After Bakhtin: Essays on Fiction and Criticism* (Routledge 1990) 96.
[30] Ward (n 3) 56.

high politics'.[31] Falstaff's world is certainly entertaining. For instance, Falstaff's drunken exaggeration of his defence during the robbery is comedic; he begins by describing how he fought off the robbers; 'sixteen at least, my lord'.[32] This number, and Falstaff's fictional bravery is soon stretched to a valiant defence against 'fifty of them'.[33] Shakespeare may have written Falstaff simply as an entertaining character; he was so popular that 'a further play was written about him [The Merry Wives of Windsor] just to please Queen Elizabeth the first'.[34] However, this world of madness could also be read as 'politically and socially subversive'.[35]

It may be that Shakespeare's Eastcheap is a commentary of what England could be without the leadership qualities learned by prince Hal. This is an England without effective governance or effective laws. For instance, in act four, scene two we see Falstaff's corrupt leadership. In enlisting soldiers for battle he 'misused the King's press damnably'[36] by accepting bribes he has got, 'in exchange of a hundred and fifty soldiers, three hundred and odd pounds'.[37] Falstaff's drunken behaviour and shambolic leadership could be analysed in the context of another Bahktin's theme, that of the carnivalesque.

This is the 'force which illustrates the way the principles of inversion and permutation work underneath the surface of carnival and festive misrule'.[38] The sense of carnival pervades scenes of Eastcheap, where drunkenness and humour undermine authority, Falstaff's status as a knight is laughable given his perpetual inebriation. Knowles suggests that even the name of the tavern, *The Boars Head*, is carnivalesque as it represents 'the main sustenance and means of sport of the big-hearted rebels'.[39] He sees the depiction of Eastcheap's revelry 'as a distorting mirror to reflect and undermine the upper level of court life and of the law'.[40] However, Knowles goes too far; whilst the carnival does break down hierarchy in society, the world of Eastcheap retains autonomous in the play and its carnival behaviour remains internal. If it is intended as a political comment, it does so by presenting dysfunctional alternative society. Again, this shows how Shakespeare presents the elements of a discussion to the audience, and allows them to form their own decision.

Shakespeare contrasts the misrule of Eastcheap with the education and subsequent success of prince Hal. He shows that without the successful performance of the princely role, a lacuna of authority is formed. This occurs in Eastcheap, where 'the idea of enforcing order by the portrayal of disorder was itself contradictory and to a degree

[31] Ward (n 1) 68.
[32] Shakespeare (n 13) 2.4.167.
[33] ibid 2.4.178.
[34] Freeman (n 17) 5.
[35] Ward (n 1) 68.
[36] Shakespeare (n 13) 4.2.11.
[37] ibid 4.2.12-13.
[38] Ronald Knowles, *Shakespeare and the Carnival: After Bakhtin* (Palgrave Macmillan 1998) 83.
[39] ibid 86.
[40] ibid 83.

destabilising'.[41] The art of effective and lawful governance is required to hold society and its rules together, as John Locke states: 'where-ever law ends, tyranny begins'.[42] As far as the law is concerned, Shakespeare is well known for poking fun at the profession. In *Henry VI Part 2* his character, Dick the Butcher, famously said 'the first thing we do, let's kill all the lawyers'.[43] This illustrates lawyer's unpopularity; in this instance, because of 'the way lawyers write, which the rebel Jack Cade insists can "undo a man"'.[44] However, society needs the law, as Llewellyn suggests, 'lawyers are unpopular and often criticised ... because they protect essential social values and roles'.[45] Shakespeare also reflects this; the world of misrule highlights the importance of law despite its unpopularity.

The involvement of law is central to the debate that Shakespeare presents upon the nature of governance. This debate is best accessed in *Richard II* where the audience is presented with the issue of whether Richard should be subject to law or not. Again, it is impossible to say whether or not Shakespeare is making a political argument here; '*Richard II* should not be read as supporting royal absolutism, or denying it, but rather as offering a way of thinking about it'.[46] This is primarily a theatrical performance; however, it displays the example of potentially destructive leadership. Even so, parallels can be drawn between *Richard II* and the carnivalesque world of Eastcheap. Richard's control over England as Falstaff's control over Eastcheap, is unstable. Shakespeare's Richard literally lost financial control of England by '"farming" out his own realm'.[47] In turn this may be a commentary upon the instability towards the end of Elizabeth's reign, such as her 'failure to raise revenue [which] led to an increasing complaint of "lack of governance"'.[48] However, Richard's 'lack of governance' also considers whether or not a king should be subject to law. For instance, Richard prevents the legal settlement of a duel between Mowbray and Bolingbroke by ordering the duellists to 'return back to their chairs again, withdraw with us; and let the trumpets sound'.[49] Where Traversi sees Richard as being 'thoroughly respectable in his motive ... following his duty',[50] a more legally minded critic may see Richard as using his royal prerogative powers far too extensively, and acting above the common law.

Richard's actions certainly break the rule of law, which requires that 'all persons and authorities within the state, whether public or private, should be bound by and entitled

[41] Ward (n 3) 204.
[42] John Locke, *Second Treatise of Government* (first published 1690, Cambridge University Press 1988) 400.
[43] William Shakespeare, 'Henry VI Part 2' in Peter Alexander (ed), *William Shakespeare: The Complete Works* (first published 1951, Collins 1985) 4.2.
[44] Weisberg (n 9) 216.
[45] Karl N Llewellyn, *The Bramble Bush: The Classic Lectures on the Law and Law School* (first published 1930, Oxford University Publishing 2008) 156.
[46] Boyd White (n 2) 47.
[47] Ward (n 1) 82.
[48] ibid 60.
[49] William Shakespeare, 'Richard II' in Peter Alexander (ed), *William Shakespeare: The Complete Works* (first published 1951, Collins 1985) 1.3.120-121.
[50] Traversi (n 12) 17.

to the benefit of laws publically made'.[51] Yet, Shakespeare once again lets the audience determine what good governance is. On one side, Richard – strictly speaking – legally implemented his royal prerogative. By contrast, interfering with justice is 'the most serious abrogation of monarchical responsibility'.[52] In the case of the latter, Richard II fails to distinguish between power and authority, and misunderstands that the 'mere temporary ascendancy of one person over another is naturally thought of as the polar opposite of law'.[53] This suggests that Richard has misinterpreted the art of governance. His public persona, unlike Hal's, is self-indulgent and power hungry. Where Hal left the lawless Eastcheap behind to pursue good leadership, Richard only learnt the importance of good counsel, public performance and being subject to the law when it was all too late.

It is only when Richard has lost everything that he sees the effect of not maintaining distance between the private self and the public persona. Then, in act four, scene one it becomes apparent that 'after the deposition there is no King, but only a man'.[54] Shakespeare uses a snow metaphor to illustrate the fragility of the position of the monarch; 'O that I were a mockery king of snow, standing before the sun of Bolingbroke, to melt myself away in water-drops!'[55] In this dramatic scene the audience is shown that behind the title of King there is a man, as fragile and subject to the law as any other. Any public performance of any ruler should therefore reflect this status.

Again, Shakespeare presents these ideas not as instruction in how to govern, but rather as a dramatisation of a poor leader. This dramatisation, however, was relevant enough to the society, for Queen Elizabeth the first to state 'I am Richard II, know ye not that?'[56] *Richard II* can be drawn upon as an informative comment upon the nature of governance and the debate about rulers treading the fine line between authority, power and being subject to law. This is a prominent theme in Shakespeare's history plays; even Hal's legal conduct is questionable in *Henry V*. His decision to go to war with France is disputably illegal as 'during the Agincourt campaign in 1415 ... a body of customary international law began to grow up'.[57] The question of the monarch's exercise of power was a contemporary debate as Shakespeare was writing. For example, the future King of England, James I's 1598 work on absolutism, *The Trew Law of Free Monarchies*, 'did little to calm sensitive common law consciences'.[58] Shakespeare's works can be considered part of this debate.

This article has shown that Shakespeare was an important voice during a period of great uncertainty in society. His characters highlight the responsibilities of royalty from a

[51] Thomas Henry Bingham, *The Rule of Law* (Penguin 2010) 8.
[52] Ward (n 1) 81.
[53] HLA Hart, *The Concept of Law* (first published 1961, Clarendon 1997) 24.
[54] Boyd White (n 2) 75.
[55] Shakespeare (n 49) 4.1.260-263.
[56] Boyd White (n 2) 50.
[57] Bingham (n 51) 30.
[58] Ward (n 1) 65.

theatrical context. Prince Hal depicts the importance of a princely education and the necessity for rulers to perform a public role if they are to lead effectively. By contrast, Shakespeare's depiction of Eastcheap illustrates a version of England without effective leadership. Shakespeare also underlines the importance of the monarch acting as subject to the law in both public and private bodies; he uses Richard II to depict a ruler who has misjudged this art of governance. Shakespeare's characters were relevant to society as examples of both effective and ineffective leadership. His ability to capture England's anxieties and reflect them in theatrical production is important both as a contribution to dramatic performance and as a comment upon early modern society.

THE ANGLO-GERMANIC BOARD ARCHITECTURE DEBATE: AN HISTORICAL AND PHILOSOPHICAL ANALYSIS

*TOM HAMILTON**

This article seeks to provide a modest contribution to the Anglo-Germanic board architectures debate as it argues that the existing debate, conducted in black-letter terms, lacks conceptual analysis. The unresolved tension between the competing forms is a result, not merely of divergent legal traditions but of major historical and philosophical differences. In order to achieve this, this paper develops three strands of intellectual genealogy and applies them to the debate. The first insight places the board architectures in the context of two competing philosophical traditions which were prominent in England and Germany during the formative years of company law: utilitarianism and Kantianism. It is submitted that the structural differences are better understood when considered in conjunction with the conception of the individual each philosophy produces. It also demonstrates that the board architectures are built on different philosophical ideas about the treatment of the individual. The second insight develops the analysis further: it argues that the corporate objectives of English and German companies are different. The third insight moves wider as it argues that England and Germany have fundamentally different market ideologies, or dominant conceptions of capitalism. The paper concludes by drawing out the implications of this genealogy for the contemporary convergence debate.

1 INTRODUCTION

Board architecture concerns the formal structure of the board of directors, the company's ultimate decision-making body.[1] As corporate governance is the 'institutional balancing process whereby the sometimes conflicting interests of a corporation's stakeholders ... are accounted for and prioritised in order to produce benefit for society',[2] the institutional design of the board has received considerable academic attention.[3] Thus far, commentators seem to take one of two approaches: an economic analysis or discussion of the differences and their supposed advantages or disadvantages.[4] The debate is reflected in the Statute of the European Public Limited-

* Tom Hamilton, Newcastle University, LLM International Business Law.

[1] Bob Tricker, *Corporate Governance: Principles, Policies and Practice* (OUP 2009) 61.

[2] Alan Dignam and Michael Galanis, *The Globalization of Corporate Governance* (Ashgate 2009) xi.

[3] Bob Tricker (n 1); Christine A Mallin, *Corporate Governance* (3rd edn, OUP 2010); Reiner Kraakman, John Armour, Paul Davies, Luca Enriques, Henry Hansmann, Gerard Hertig, Klaus Hopt, Hideki Kanda and Edward Rock (eds), *The Anatomy of Corporate Law: A Comparative and Functional Approach* (2nd edn, OUP 2009); Michael Aglietta and Antoine Rebérioux, *Corporate Governance Adrift: A Critique of Shareholder Value* (Edward Elgar 2005).

[4] Petri Mäntysaari, *Comparative Corporate Governance: Shareholders as a Rule Maker* (Springer 2005); Bill Perry and Lynne Gregory, 'The European Panorama: Directors' Economic and Social Responsibilities' [2009] ICCLR 25; Florian Schwarz, 'The German Co-Determination System: A Model for Introducing Corporate Social Responsibility Requirements into Australian Law (Part 1)' [2008] Journal of International Banking Law and Regulation 125; Florian Schwarz, 'The German Co-Determination System: A Model for Introducing Corporate Social Responsibility Requirements into Australian Law (Part 2)' [2008] Journal of International Banking Law and Regulation 190; Gavin Kelly

Liability Company, or *Societas Europaea* (SE), in so far as the SE provides companies with a limited choice between the unitary, or 'single-tier' board favoured in England and the two-tier board mandated by German law.[5]

English practice comprises a unitary board composed of executive and non-executive directors.[6] Non-executive directors may be connected to the company in some way or completely independent (INEDs).[7] All directors are under a statutory duty to promote the success of the company and are not, therefore, considered to be delegates of any particular stakeholder constituency.[8] In theory, the independent directors supervise the executive directors in the running of the company and in so doing protect shareholders and other corporate constituents against potential abuses or excesses of the executive directors.[9]

In German law large enterprises, those employing over 2,000 permanent employees, regardless of legal form, are bound by board-level codetermination.[10] Such companies are required to have a two-tier board structure comprising the *vorstand*, or management board, composed entirely of executive directors and the *aufsichtsrat*, or supervisory board, which comprises representatives of workers and shareholders in equal amounts.[11] The casting vote is assigned to the chairman who is always a shareholder representative and prevents deadlock.[12] The *aufsichtsrat*, in theory, monitors and approves the *vorstand*'s decisions and enjoys powers of appointment and dismissal over it.[13] Indeed codetermination reflects the German perception of enterprise as a coalition of capital and labour, the classical factors of production.[14] An explicit duty to promote shareholder value is not found in German law.[15]

---

and John Parkinson, 'The Conceptual Foundations of the Company: a Pluralist Approach' in John Parkinson, Gavin Kelly and Andrew Gamble (eds), *The Political Economy of the Company* (Hart 2001); Horst Siebert, 'Corporatist verses Market Approaches to Governance' in Klaus J Hopt, Eddy Wymeersch, Hideka Kanda and Harald Baum (eds), *Corporate Governance in Context: Corporations, States and Markets in Europe, Japan and the US* (OUP 2005).

[5] Council Regulation 2001/2157/EC of 8 October 2001 on the Statute for a European company (SE) [2001] OJ L 294/1; Council Directive 2001/86/CE 8 October 2001 supplementing the Statute for a European company with regard to the involvement of employees [2001] OJ L 294/22.

[6] Bob Tricker (n1) 64 – 67.

[7] Bob Tricker (n1) 50 – 51.

[8] Companies Act 2006, s 172(1); Bill Perry and Lynne Gregory (n4) 27-9; cf Andrew Keay, 'Moving Towards Stakeholderism? Constituency Statutes, Enlightened Shareholder Value, And All That: Much Ado About Little?' (2011) 22(1) European Business Law Review 1; Lisa Linklater, 'Promoting success: the Companies Act 2006' (2007) 17 Comp Law 129.

[9] Bob Tricker (n1) 63; cf Petri Mäntysaari (n4) 401-4.

[10] Harald Baum, 'Change of Governance in Historic Perspective: The German Experience' in Klaus J Hopt, Eddy Wymeersch, Hideka Kanda and Harald Baum (eds), *Corporate Governance in Context: Corporations, States and Markets in Europe, Japan and the US* (OUP 2005) 16; Manfred Balz, 'Corporate Governance in Germany' (OECD Conference, Seoul, March 1999); Horst Siebert (n4) 287.

[11] Horst Siebert (n4) 287-8; Michel Aglietta and Antoine Rebérioux (n3) 57.

[12] Horst Siebert (n4) 287-8.

[13] Petri Mäntysaari (n4) 255, 263; Christine A Mallin, *Corporate Governance* (3rd edn, OUP 1999) 215; Florian Schwarz (Part 2) (n 4) 190-2.

[14] Manfred Balz (n10) 2; cf Harald Baum (n10) 9; Horst Siebert (n4) 289.

[15] Alan Dignam and Michael Galanis (n2) 274.

This paper attempts to address two distinct lines of inquiry. Firstly, it seeks to elucidate the historical and intellectual reasons for divergence between English and German board architectures. Secondly, it seeks to address the possibility of accommodation between the two positions in the context of the European Union (EU) harmonisation project.

1.1 Tacit Assumptions

The debate between proponents of the English and German structures is of particular interest and relevance given the emergence of the European corporate form. The *Societas Europaea* (SE) embodies an awkward compromise in the harmonisation of corporate governance within the EU.[16] This debate has been largely conducted in black-letter terms and concerned with the discussion of the relative advantages and disadvantages of the competing architectures.[17] As such, the author hopes to make a modest contribution to this debate by deepening its conceptual analysis. It is submitted that it has fallen into the trap in which Professors Hart and Fuller found themselves over fifty years ago.[18] In the words of Professor Fuller:

> As critical reviews of my book came in, I myself became increasingly aware of the extent to which the debate did indeed depend on "starting points"—not on what the disputants said, but on what they considered it unnecessary to say, not on articulated principles but on tacit assumptions. What was needed therefore, it seemed to me, was to bring these tacit assumptions to more adequate expression than either side has so far been able to do.[19]

Without an adequate exploration of starting points the debate has, it is submitted, fallen prey to the trap set in the Mad Hatter's Riddle: "[w]hy is a raven like a writing-desk?"[20] The reader, like Alice, presupposes that there is both similarity and solution and proceeds based upon that assumption. Therefore this paper is a work of intellectual genealogy and wishes to bring to light previously buried assumptions or 'starting points' in the hope that the debate may move beyond its present impasse.

As such, this paper will proceed based upon the observation that the existing debate, conducted in black-letter terms, is not merely artificial but has grown stale, stifled by a neglect of conceptual analysis.[21] It will then argue that the unresolved tension between the competing forms is a result, not merely of divergent legal traditions but, of major historical and philosophical differences. This thesis shall be demonstrated by the

---

[16] Statute for a European company (n5).
[17] Petri Mäntysaari (n4); Bill Perry and Lynne Gregory (n4); Florian Schwarz (n4).
[18] HLA Hart, *The Concept of Law* (2nd edn, OUP 1994); Lon L Fuller, *The Morality of Law* (Revised edn, Yale University Press 1969).
[19] Lon L Fuller (n18) 189; cf HLA Hart, 'Book Review – The Morality of Law' (1965) 78 Harvard Law Review 1281.
[20] Lewis Carroll, *Alice's Adventures in Wonderland* (MacMillan 1920) 97.
[21] Petri Mäntysaari (n4); Bill Perry and Lynne Gregory (n4); Florian Schwarz (n4).

application of three exegetical tools to the current debate. The extent to which these are convincingly established will determine the validity of the thesis.

At this juncture, it is appropriate to address the tacit assumptions, or methodological considerations, upon which the paper will be based. Law is not a hermeneutically sealed discipline.[22] It is conceptually open and develops alongside the political, philosophical, moral and economic fabric of society.[23] These influences, however, are not always clear or acknowledged and are often buried below legal verbiage. As such, English and German economic history, as well as law, will be deployed to demonstrate the philosophical insights: this is the second respect in which this paper can be considered a work of intellectual genealogy. If Thomas Reed Powell's caricature of the 'legal mind' as 'the ability to think about something which is attached to something else without thinking about what it is attached to' is accurate, this paper will be unrecognisable as the product of such a mind.[24]

1.2 Three Exegetical Tools

Following a review of the literature in the following section, this paper shall develop three strands of intellectual genealogy and apply them to the debate. In section three, the first insight shall be considered. It places the board architectures in the context of two competing philosophical traditions which were prominent in England and Germany during the formative years of company law: utilitarianism and Kantianism.[25] It will be submitted that the structural differences are better understood when considered in conjunction with the conception of the individual each philosophy produces. It will be demonstrated that English and German board architectures are built on different philosophical ideas about the treatment of the individual.

In section four, the second insight develops the analysis further: it will be argued that the corporate objectives of English and German companies are different. Proponents seem tacitly to assume that the underlying purpose of companies in both England and Germany is the same.[26] In so doing, they do not merely fall into the Hatter's trap[27] but commit what Nietzsche called 'the commonest act of stupidity': 'forgetting one's purpose.'[28] It is argued that the English commitment to shareholder value[29] is radically

[22] John Finnis, *Natural Law and Natural Rights* (Oxford 1980) 14-8.
[23] David J Gerber, *Law and Competition in Twentieth Century Europe: Protecting Prometheus* (OUP 1998) ix, x, xxi.
[24] Donald L Horowitz, 'The Cracked Foundations of the Right to Secede' (2003) 14(2) Journal of Democracy <http://www.journalofdemocracy.org/articles/gratis/Horowitz.pdf> Accessed 10 February 2012, 4; cf Thurman Arnold, 'Criminal Attempts—The Rise and Fall of Abstraction' (1930-31) 40 Yale Law Journal 58.
[25] Alan Dignam and Michael Galanis (n2) 263 – 392.
[26] Petri Mäntysaari (n4); Bill Perry and Lynne Gregory (n4); Florian Schwarz (n4).
[27] Lewis Carroll (n20).
[28] Friedrich Nietzsche, *Menschliches, Allzumenschliches: ein Buch für freie Geister*, vol 2 (1879) para 206; cf Lon L Fuller (n18) 95.
[29] Andrew Keay (n8).

different from the German view which is allied to stakeholder theory and results in a bifocated corporate objective and this difference informs board architecture. [30]
The third insight, developed in section six, moves wider again. The argument is made that England and Germany have fundamentally different market ideologies, or dominant conceptions of capitalism. These different conceptions of capitalism are identified before being traced into the legal structures that emerged.[31] The paper concludes in section six by drawing out the implications of this genealogy for the contemporary convergence debate.

1.3 Critical Scope

The paper will be limited to comparative discussion of board architecture in England and Germany since the eighteenth century. A number of other countries such as the Netherlands, France, Japan, Korea and China have distinctive and interesting board architectures and traditions, the consideration of which would enrich any comparative analysis. Their exclusion, although regrettable, is necessary.[32] Not the least for considerations of space: in order to develop a sufficiently rigorous conceptual analysis the field has been restricted.

The selection of England and Germany is based on the scope of debate in European circles, particularly within the Commission reports, which has been conducted between these positions.[33] Indeed, so pressing did the EU find this debate that it carried the controversy through into the SE.[34] Tellingly, many continental systems incorporate features of both English and German forms to some extent.[35] Therefore, in drawing the comparison between these two forms we find not only the sharpest distinction and the most academic controversy but greatest legal relevance. A notable distinction is also found in the philosophical and intellectual traditions of England and Germany; the tension between Kantian and utilitarian ethics is heavily discussed in legal theory.[36] Therefore, the choice of comparison is not merely intellectually preferable, but most politically relevant.

---

[30] Gavin Kelly and John Parkinson (n4) 130-1.
[31] Karl Marx, *The German Ideology* (International Publishers 1939) 18.
[32] Jin Zhu Yang, 'The Anatomy of Boards of Directors: An Empirical Comparison of UK and Chinese Corporate Governance Practices' (2007) Company Law 24; John Buchanan and Simon Deakin, 'Japan's Paradoxical Response to the New 'Global Standard' in Corporate Governance' ECGI Law Working Paper No. 87/2007 <http://ssrn.com/abstract=1013286> accessed 14 March 2012; Bob Tricker (n1) 67.
[33] Klaus J Hopt, 'The German Two-Tier Board: Experience, Theories, Reforms' in Klaus J Hopt, Hideki Kanda, Mark J Roe, Eddy Wymeersch and Stefan Prigge (eds), *Comparative Corporate Governance: The State of the Art and Emerging Research* (Clarendon Press 1998) 228; Bob Tricker (n1) 67.
[34] Statute for a European company (n5).
[35] Klaus J Hopt (n33) 228-9; Bob Tricker (n1) 67.
[36] John Rawls, *A Theory of Justice* (Harvard University Press 1971).

2 LITERATURE REVIEW

The board architecture debate has been colourful: codetermination has been cast as both the cause of German economic stagnation in the 1990s and early 2000s and a solution to the governance problems in Anglo-Saxon firms that lead to the credit crunch.[37] It is argued that the debate, despite its colour and passion, has been stifled by neglect of conceptual analysis; it hinges on hitherto unarticulated starting points or 'tacit assumptions'.[38] This paper therefore is an attempt to 'bring to more adequate expression' the tacit assumptions that underpin both systems.[39] These tacit assumptions concern: fundamentally different philosophical conceptions of the individual,[40] different corporate objectives[41] and different market ideologies.[42] It is argued that these divergent assumptions explain the differences of legal structure and the intractable nature of the debate.

This literature review shall demonstrate that the debate has been conducted largely in terms of advantages and disadvantages and economic analysis and therefore it cannot address the question of why these differences persist or whether there is, truly, any room for accommodation between the two positions at the level of principle, within the harmonisation project.

2.1 Advantages and Disadvantages

Much of the literature focuses on the comparative advantages and disadvantages of the various board architectures and discusses the structural differences technically and in detail. Advocates of codetermination contend that it has helped to prevent abuses of the management-dominated boards that have occurred in the unitary system.[43] The controls on executive power are underlined by the absence of a single managing director (MD) although the chairman of the *vorstand* may in practice act as a MD there is no necessary concentration of executive power in a single individual; instead the *vorstand* operates as a collective organ.[44] A strength of codetermination is the representation of a number of stakeholder interests, such as employees and creditors, alongside shareholders on the

[37] Harald Baum (n10) 16; Petri Mäntysaari (n4); Horst Siebert (n4); Alan Dignam and Michael Galanis (n2) xix, 45-6.
[38] Lon L Fuller (n18) 189; cf HLA Hart (n19).
[39] ibid.
[40] Immanuel Kant, *Groundwork for the Metaphysics of Morals* (Allen W Wood tr, Yale University Press 2002) 47; John Stuart Mill, *Utilitarianism and the 1868 Speech on Capital Punishment* (George Sher Ed, 2nd edn, Hackett 2001) 7.
[41] eg Andrew Keay (n8); Michel Aglietta and Antoine Rebérioux (n3).
[42] Karl Marx, *The German Ideology* (International Publishers 1939); Friedrich August von Hayek, *The Constitution of Liberty* (Routledge 2006); Friedrich August von Hayek, *The Road to Serfdom* (Routledge 2001); David J Gerber, 'Constitutionalising the Economy: German Neo-liberalism, Competition Law and the "New" Europe' (1994) 42 American Journal of Comparative Law 25; Massimiliano Vatiero, 'The Ordoliberal Notion of Market Power: An Institutionalist Reassessment' (2010) 6(3) European Competition Journal 689; Viktor J Vanberg, 'The Freiburg School: Walter Eucken and Ordoliberalism' (2004) Freiburg Discussion Papers on Constitutional Economics 04/11 <https://www.econstor.eu/dspace/bitstream/10419/4343/1/04_11bw.pdf> accessed on 14 March 2012.
[43] Thomas Clarke and Richard Bostock, 'Governance in Germany: The Foundations of Corporate Structure' in Kevin Keasey, Steve Thompson and Mike Wright (eds), *Corporate Governance: Economic, Management and Financial Issues* (OUP 1997) 245; Petri Mäntysaari (n4) 252.
[44] Petri Mäntysaari (n4) 259-61.

aufsichtsrat.[45] Clarke and Bostock note the shareholder and worker representation increases accountability and reduces institutional pressures towards short-term decision-making[46] although the historically less developed German capital market and important role of house banks are also credited.[47] A number of social benefits are accepted as resulting from such a system, which include improved labour relations and cooperation within the organisation[48] and recognition of employees' intelligence and skill as principal assets of the business.[49] However, the negative form of long-termism is rigidity: codetermination makes it more difficult for companies to change quickly or take painful measures such as redundancies.[50]

The independence of the *aufsichtsrat* is underpinned by a strict separation of responsibilities, powers and personnel between the tiers.[51] The *aufsichtsrat* has control over its own appointment and there is no formal process for the *vorstand* or its members to make nominations.[52] Mäntysaari, observing the more-or-less automatic appointment of the outgoing *vorstand* chairman to the chairmanship of the *aufsichtsrat*, asserts that such close contacts, rather than undermining the independence of the *aufsichtsrat*, enable it to take care of its supervisory and advisory role.[53] However, he also notes that lack of independence in the *aufsichtsrat* was perceived as a problem for German corporate governance[54] and notes that as a rule the *vorstand* has an exclusive right within the company to disclose information to the *aufsichtsrat* regarding management matters.[55]

More critically, Clarke and Bostock note this information reliance and point out this dialogue has failed to prevent corporate crises such as the (near) total collapses of Metallgesellschaft, Daimler-Benz, Klocker-Humbolt-Deutz, Holzmann and Kirch.[56] *Aufsichtsrate* meet rarely, for example Volkswagen's met only four times per year, twice more than then statutory minimum.[57] Clarke and Bostock contrast this with the unitary board which will often meet monthly, perhaps more frequently in sub-committees[58] and will therefore normally have a quicker reaction time.[59] They also note

[45] Petri Mäntysaari (n4) 261; Florian Schwarz (n4) 128ff; Thomas Clarke and Richard Bostock (n43) 244.
[46] Thomas Clarke and Richard Bostock (n43) 244.
[47] Michel Aglietta and Antoine Rebérioux (n3) 49-57; Thomas Clarke and Richard Bostock (n43) 235-44; cf Horst Siebert (n4) 291.
[48] Thomas Clarke and Richard Bostock (n43) 245.
[49] ibid 246.
[50] Horst Siebert (n4) 290; Thomas Clarke and Richard Bostock (n43) 246.
[51] Petri Mäntysaari (n4) 260 – 261; Florian Schwarz (4) 190 – 193.
[52] Petri Mäntysaari (n4) 261.
[53] ibid 262; Klaus J Hopt (n33) 233.
[54] Petri Mäntysaari (n4) 263; Florian Schwarz (Part 2) (n4) 192.
[55] Petri Mäntysaari (n4) 266; Florian Schwarz (Part 2) (n4) 192; Alan Dignam and Michael Galanis (n2) 292.
[56] Thomas Clarke and Richard Bostock (n43) 243; Alan Dignam and Michael Galanis (n2) 342.
[57] ibid 243.
[58] Thomas Clarke and Richard Bostock (n43) 243.
[59] ibid 246.

that INEDs enjoy greater access to information than the *aufsichtsrat*.[60] Heinrich Weiss identified two serious flaws in the German system: firstly, the presence of labour representatives prevents shareholder representatives putting critical questions to the management for fear of undermining their authority in the presence of employees. Secondly, this has led to a situation where *vorstand* members invite friends and colleagues to join *aufsichtsrate,* thereby leading to control of the *aufsichtsrat* by the *vorstand*, rather than the other way round.[61] Similarly, Hansmann and Kraakman argue that codetermination is inefficient, likely to result in paralysis or weak boards and these costs are likely to exceed any benefit that worker participation might bring.[62] In light of this, the picture of independence and effective control painted by Mäntysaari, Schwarz and others appears somewhat less than convincing.

Mäntysaari identifies several fundamental problems within UK company law which he claims do not exist within the German system; they centre around what he perceives as a failure to separate supervision and management at board-level.[63] These include wide discretionary powers on the part of the board which cannot be constrained by mere disclosure rules and 'loosely defined guidelines' made in case law.[64] Mäntysaari also points to wide-spread delegated decision-making in UK companies which he argues has led to a situation where the *de jure* board acts primarily as a supervisory organ over a management structure regulated by the internal rules of the company, creating a *de facto* two-tier governance regime.[65] He argues that this leaves the board with an excess of powers which, he assumes, need to be constrained.[66] Mäntysaari points to requirements in legislation and the Combined Code which require a separation between independent directors and executive directors and creation of audit committees.[67] He argues that the existing statutory controls are insufficient as they do not state how the powers of the board should be exercised and that external rules need to be imposed to govern the exercise of these powers, and those of the sub-board level managers who actually run the company.[68] For Mäntysaari, such problems have not arisen in German law as it prohibits the delegation of management by the *vorstand* and there is considerable constraint on managers by virtue of the *aufsichtsrat* and prescriptive statutory provisions.[69] This prohibition of delegation means that the statutory controls lacking in English law are unnecessary in German law.[70] However, Prigge, drawing on some

[60] ibid 243.
[61] ibid 243 – 244; Klaus J Hopt (n33) 231-2.
[62] Henry Hansmann and Reinier Kraakman, 'The End of History For Corporate Law' in Jeffrey N Gordon and Mark J Roe (eds), *Convergence and Persistence in Corporate Governance* (CUP 2004) 33, 52, 57
[63] Petri Mäntysaari (n4) 401, 403.
[64] ibid 401.
[65] ibid 401.
[66] ibid 401.
[67] ibid 402; cf Companies Act 2006.
[68] ibid 402.
[69] ibid 403.
[70] ibid 403.

empirical evidence, suggests that *aufsichtsrate* codetermination actually weakens the intensity, and therefore effectiveness, of supervisory control.[71]

Mäntysaari argues that the flexibility of the English structure requires investors to analyse the governance structure of each company separately and that the prescriptive, and standardised, nature of German governance guarantees the constraints on executive powers that are essential for shareholder protection.[72] Mäntysaari asserts the INEDs are incapable of protecting shareholders' interests without detailed and binding rules against which to measure executive conduct.[73] Baums and Scott, however, take a different approach and make several criticisms of German law. They claim it inadequately regulates for: conflicts of interest between controlling shareholders, directors and the corporation; the inability to use derivative actions by shareholders to police bad performance; breaches of duty of care; and an inability to enforce the informational obligations owed to investors.[74]

2.2 Economic Analysis

A number of authors contribute to the debate by conducting an economic analysis.[75] Hansmann and Kraakman argue that codetermination is economically inefficient in contrast to the unitary board.[76] Siebert paints codetermination as a symptom of Germany's restrained market economy[77] and argues that the present system protects the employed (insider) to the detriment of the unemployed (outsider)[78] and that this problem can only be solved by movement towards a free market,[79] which necessarily involves weakening the institutional representation of labour on the board.[80] Further, Siebert points to the conflict of interest for labour representatives between the interests of the firm and those of the union.[81] Like Hansman and Kraakman, Siebert considers codetermination to be untenable under conditions of globalisation.[82] They argue that board architecture is converging in the direction of the unitary board, as part of a wider trend towards convergence away from a stakeholder model to one of shareholder

[71] Stefan Prigge, 'A Survey of German Corporate Governance' in Klaus J Hopt, Hideki Kanda, Mark J Roe, Eddy Wymeersch and Stefan Prigge (eds), *Comparative Corporate Governance: The State of the Art and Emerging Research* (Clarendon Press 1998) 1010.
[72] Petri Mäntysaari (n4) 404.
[73] ibid 405.
[74] Theodor Baums and Kenneth E Scott, 'Taking Shareholder Protection Seriously? Corporate Governance in the United States and Germany' ECGI Working Paper No. 17/2003 <http://ssrn.com/abstract=473185> accessed 14 March 2012, 19.
[75] ibid; Klaus J Hopt (n33) 238ff.
[76] Henry Hansmann and Reinier Kraakman (n62) 45-8.
[77] Horst Siebert (n4) 281.
[78] ibid 290.
[79] ibid 286.
[80] ibid 286 – 289.
[81] Horst Siebert (n4) 289; cf Klaus J Hopt (n33) 237.
[82] ibid 292.

primacy.[83] In making this argument, they point to the failed EC Fifth Directive[84] on Company Law which attempted to extend codetermination throughout Europe, the fact that enthusiasm for codetermination has receded to the point of an option in the SE and the liberalisation of capital markets.[85] Mäntysaari and Davies argue that convergence is actually occurring in the opposite direction pointing towards the use of supervisory functions of board sub-committees in English companies, for example audit and remuneration.[86]

2.3 Divergent Views of Convergence

Conversely, Aglietta and Rebérioux advocate codetermination as a crucial aspect of economic democracy, as opposed to shareholder value.[87] Rather than seeing the board of directors in terms of economic agency and contract, they see it as a political agent charged with balancing the concerns of various stakeholders in the public interest.[88] This, they argue, is the preferable form of corporate governance because capitalism cannot foster social progress if the market is not subject to democratic control.[89] However, the notion of social progress is philosophically questionable as is the propriety of corporate governance as a vehicle to foster it.[90]

Dignam and Galanis consider the debate in terms of convergence and draw a distinction between insider systems such as the German and outsider systems such as the Anglo-Saxon.[91] They typify codetermination as cooperative rather than combative,[92] and point to the low political appetite for challenging codetermination.[93] Like Hansman and Kraakman, Dignam and Galanis are equivocal about the future of codetermination, citing the much used freedom of companies to choose their state of incorporation since *Centros Ltd v Erhvervs-og Selskabsstryrelsen*[94] and the possibility to incorporate as an SE with a unitary board under EU law.[95] They point to the popularity of incorporation

---

[83] Henry Hansmann and Reinier Kraakman (n62).
[84] Proposal for a Fifth Company Law Directive founded on article 54(3)(g) of the EEC Treaty concerning the structure of public limited companies and the powers and obligations of their organs, 1972 EC Bull Supp 10 ('1972 Proposed Fifth Directive').
[85] ibid 5, 18-9.
[86] Petri Mäntysaari (n4) 400; Paul Davies, 'Board Structure in the UK: Convergance or Continuing Divergence?' Working Paper <http://ssrn.com/abstract=262959> accessed 14 March 2012, 2.
[87] Michel Aglietta and Antoine Rebérioux (n3) 266.
[88] ibid 267.
[89] ibid 273.
[90] Michael Oakshott, 'Rationalism in Politics' in Timothy Fuller (ed), *Rationalism in Politics and Other Essays* (Liberty Fund 1991); John Gray, *Straw Dogs: Thoughts on Humans and Other Animals* (Granta 2002); John Gray, *Gray's Anatomy* (Allen Lane 2009).
[91] Alan Dignam and Michael Galanis (n2) 43 – 44.
[92] ibid 267.
[93] ibid 312-5, 342-3.
[94] (Case C-212/97) [1999] ECR 1459; Jens C Dammann, 'The Future of Coodetermination After *Centros*: Will German Corporate Law Move Closer to the US Model?' 2003 Financial Law 607.
[95] Council Directive 2001/86/CE 8 October 2001 supplementing the Statute for a European company with regard to the involvement of employees [2001] OJ L 294/22; Council Regulation 2001/2157/EC of 8 October 2001 on the Statute for a European company (SE) [2001] OJ L 294/1; Alan Dignam and Michael Galanis (n2) 315 – 319; Jochem Reichert, 'Experience with the SE in Germany' (2008) 4(1) Utrecht Law Review 22.

of German business under English law citing the belief that codetermination 'is no longer competitive internationally'.[96] This does not necessarily signal the triumph of the unitary board, but, they suggest, may simply show that the current model is under renegotiation.[97] Baums and Scott advise the decision between architectures should be opened to competition and left to the market.[98]

2.4 Deficiencies in the Debate

This review has outlined the advantages and disadvantages of the English and German systems as well as several insights from economic analysis. However neither body of knowledge has addressed the intellectual genealogy of the debate: the basic ideas about the individual and how he ought to be accommodated within board architecture. While methods of achieving convergence have been identified, the question of whether the product of convergence could accommodate divergence at the level of principle is yet to be seriously addressed.

3 CONCEPTIONS OF THE INDIVIDUAL

English and German board architectures rest on fundamentally different tacit assumptions about the role of the individual, or worker, within the firm. This section shall articulate these philosophical assumptions and demonstrate that the board architectures, and differences between them, are better understood in this context. It will be argued that the intractable nature of the debate is a result of conflict at the level of principle as the English and German board structures are predicated opposing views of the treatment of the hypothetical individual within the corporation.

These assumptions are derived from utilitarianism and Kantianism which were prominent in the intellectual *zeitgeister* of England and Germany during the formative years of company law. It is not submitted that English and German company law, or corporate governance, are perfect vindications of these philosophies. Rather, the contention is that the ideas and patterns of thought articulated in these, and other, philosophies were influential in English and German society and corporate governance. Therefore, the analysis of corporate governance through such ideas may provide insight in the board architecture debate.[99]

3.1 Utilitarianism and the Unitary Board

Utilitarianism is a 'political creed'[100] that was first advocated in England by Jeremy Bentham,[101] in 1776,[102] and reached its most authoritative and influential articulation in

---

[96] Alan Dignam and Michael Galanis (n2) 318.
[97] ibid 320 – 321.
[98] Theodor Baums and Kenneth E Scott (n74) 19.
[99] See David J Gerber (n23) ix, x, xxi.
[100] John Stuart Mill (n40) 7.
[101] Jeremy Bentham, *An Introduction to the Principles of Morals and Legislation* (Clarendon Press 1907) available at <http://www.econlib.org/library/Bentham/bnthPMLCover.html> accessed 18 April 2013.
[102] Jeremy Bentham, *Fragment on Government* (1776 London) available at <http://www.efm.bris.ac.uk/het/bentham/government.htm> accessed 18 April 2013; JM Kelly, *A Short History of Western Legal Theory* (1992 OUP) 287.

John Stuart Mill's *Utilitarianism* of 1863.[103] It is 'the theory that actions, laws, policies and institutions are to be evaluated by their utility... by the degree to which they have better consequences than alternatives.'[104] Although it has been refined and rearticulated several times since,[105] for our purposes Mill's work is most appropriate owing to its historical importance as evidence of English Enlightenment thinking.[106] Utilitarianism was perhaps the most influential philosophical movement in British history,[107] Mill was 'the foremost British philosopher of the nineteenth century'[108] and utilitarianism was more than a mere set of ideas but the creed of a political movement of 'philosophical radicals' who published widely, held political office and sought legislative reform.[109] It continues to be widely debated not merely in academic circles but in many cases of practical ethics.[110]

Therefore, utilitarianism was a central current in the intellectual *zeitgeist* when in 1844 Parliament legislated to allow the incorporation of joint stock companies by registration, the predecessor of the modern company.[111] The 1844 Act, with subsequent amendments,[112] opened up the corporate form to a 'seven or more persons associating together and subscribing to a memorandum of association; a name and a registered office'.[113] From the outset English companies were managed by a unitary board, what Adam Smith referred to as 'a court of directors'[114] which was conceptually distinct from shareholders.[115] To form a company to conduct business soon came to be considered, in the words of Robert Lowe the then Vice-President of the Board of Trade, a 'natural right'.[116] A key feature of English company law since 1844 has been a 'collective laissez-faire',[117] or minimalist, approach to state regulation, of both company law and economic relations.[118] Therefore, the company was viewed as a private arrangement 'for

---

[103] John Stuart Mill (n40).
[104] Henry R West (ed), *The Blackwell Guide to Mill's Utilitarianism* (Blackwell 2006) 1.
[105] Henry Sidgwick, *Methods of Ethics* (7th edn, 1907 Macmillan) available at <http://www.laits.utexas.edu/poltheory/sidgwick/me/index.html> accessed 18 April 2013; JJC Smart, 'An Outline of a System of Utilitarian Ethics' in JJC Smart and Bernard Williams, *Utilitarianism: For and Against* (CUP 1963).
[106] Gerald J Postema, 'Bentham's Utilitarianism' in Henry R West (ed), *The Blackwell Guide to Mill's Utilitarianism* (Blackwell 2006) 26ff.
[107]Edward N Zalta (ed), *Stanford Encyclopaedia of Philosophy* available at <http://plato.stanford.edu/entries/utilitarianism-history/> accessed 3 August 2012.
[108] Henry R West (n104).
[109] ibid 2; JM Kelly (n102) 315-20.
[110] Henry R West (n104) 6.
[111] Joint Stock Companies Act 1844; Geoffrey Morse (ed), *Palmer's Company Law* (Sweet & Maxwell 2012) para 1.104.
[112] Joint Stock Companies Act 1856; Limited Liability Act 1856; Companies Act 1862.
[113] Susan Watson, 'The Significance of the Source of the Powers of Boards of Directors in UK Company Law' 2011(6) JBL 597, 607.
[114] Adam Smith, *An Inquiry into the Nature and Causes of the Wealth of Nations*, vol 3 (5th edn, 1789) 731
[115] Joint Stock Companies Act 1844, s 4(6), s 27-31ff; Companies Clauses Consolidation Act 1845; *Burnes v Pennell* (1849) 2 H.L. Cas. 497 HL at 521.
[116] Rob McQueen, A *Social History of Company Law: Great Britain and the Australian Colonies 1854-1920* (Ashgate 2009) 53.
[117] Otto Kahn-Freund, 'Labour Law' in Morris Ginsberg (ed), *Law and Opinion in England in the 20th Century* (University of California Press 1959) 215.
[118] Rob McQueen (n116) 19, 29, 52.

the private gain of the members as a whole'.[119] These three features: the unitary board, answerable to shareholders, pursuit of 'shareholder value' and a relatively laissez-faire regulatory approach remain central tenants of English company law.[120]
The English company has a considerable utilitarian inheritance which is reflected in its board architecture. Utilitarianism establishes ends in terms of goods, principally in the maximisation of pleasure or welfare:[121]

> The creed which accepts the foundation of morals "utility" or "the greatest happiness principle" holds that actions are right in proportion as they tend to promote happiness; wrong as they tend to produce the reverse of happiness, pain and the privation of pleasure... the theory of life on which this theory is grounded—namely that pleasure and freedom from pain are the only things desirable as ends; and that all desirable things ... are desirable either for pleasure inherent in themselves or as means to the promotion of pleasure and prevention of pain.[122]

Therefore, utilitarianism licences the employment of individuals as instruments in the pursuit of utility; as such individuals are not considered ends in themselves, but instruments to achieve more important ends. Therefore, utilitarianism can provide justification for selfishness and can legitimise the treatment of certain groups as means and the exclusion of other ends in immediate decision-making. It is submitted that this instrumentalisation of the individual is reflected in the corporate context.

By not requiring labour representation on the board, shareholders are the only constituency represented and theirs only interest served by the company.[123] The employees are merely thought of as one of a number of constituencies employed in the services of shareholder wealth maximisation, or 'shareholder value'.[124] Therefore, for English law, the end is shareholder value and employees, are a means of achieving that end, however highly trained and invested in.

---

[119] Colin Arthur Cooke, *Corporation, Trust and Company: An Essay in Legal History* (Manchester: University Press 1950) 55; Susan Watson (n113) 600.
[120] Companies Act 2006, ss 154, 168; John Armour, Henry Hansmann and Reinier Kraakman, 'Agency Problems and Legal Strategies' in Reinier Kraakman, John Armour, Paul Davies, Luca Enriques, Henry Hansmann, Gerard Hertig, Klaus Hopt, Hideki Kanda and Edward Rock,*The Anatomy of Corporate Law: A Comparative and Functional* Approach (2nd edn, OUP 2009) 35-6.
[121] John Stuart Mill (n40).
[122] ibid 7.
[123] Paul L Davies, 'A Note on Labour and Corporate Governance in the UK' in Klaus J Hopt, Hideki Kanda, Mark J Roe, Eddy Wymeersch and Stefan Prigge (eds), *Comparative Corporate Governance, The State of the Art and Emerging Research* (Clarendon Press 1998) 377; See also: Colleen A Dunlavy, 'Corporate Governance in Late 19th Century Europe and the US: The Case of Shareholder Voting Rights' in Klaus J Hopt, Hideki Kanda, Mark J Roe, Eddy Wymeersch and Stefan Prigge (eds), *Comparative Corporate Governance, The State of the Art and Emerging Research* (Clarendon Press 1998) 5.
[124] See Henry Hansmann and Reinier Kraakman (n62) 33.

3.2 Kantianism and Co-determination

Kantianism refers to the body of thought originating in the works of Immanuel Kant, 'the greatest modern philosopher'.[125] Kant's moral philosophy was organised around a standard of practical rationality he termed the 'categorical imperative'.[126] It unnecessary to précis of Kant's philosophy and historical significance, for our purposes reference to the 1785 *Groundwork for the Metaphysics of Morals* shall be sufficient owing to its prescience and influence: it is 'one of the most significant texts in the history of ethics'.[127] The influence of Kant cannot be overstated. His work has been developed, expanded and heavily critiqued by virtually every philosopher that followed and therefore stands apart as the most influential German contribution to Enlightenment thought.[128]

Therefore, Kantian ideas have been present in the German *zeitgeist* since the eighteenth century. The two-tier board structure in German company law dates back to its mandatory introduction in 1870.[129] The *vorstand* was conceptualised as the economic and legal agent of the shareholders whereas the *aufsichtsrat* began as an outside committee of shareholders which later developed into 'a unitary governing board' with representatives of shareholders, bankers and related entrepreneurs.[130] Concerns beyond the narrow paradigm of shareholder and manager were clear in the debate from the beginning. For example, Robert von Mohl expressed concerns that large corporations could corrupt legislators in their own interest at the expense of stakeholders such as labour and the public interest.[131]

Although labour codetermination did not enter German law until relatively recently, (1951 in the coal and steel industry[132] and extended generally in 1976)[133] it had first been conceived in 1835 by von Mohl and had been experimented with since 1920.[134] Its re-introduction during the denazification of West Germany was the result of a number of forces: in part a reaction to collusion between corporate Germany and the Nazi party, a recognition of workers claims for greater responsibility, a managerial means of escaping total responsibility and dissatisfaction with the results of nationalisation of coal in Britain.[135] However boardroom level codetermination emerged in the context of

[125]Allen W Wood, 'What Is Kantian Ethics' in Immanuel Kant, *Groundwork for the Metaphysics of Morals* (Allen W Wood tr, Yale University Press 2002) 157; Roger Scruton, Kant: A Very Short Introduction (OUP 1981) 1.
[126] ibid.
[127] ibid ix.
[128] ibid ix-xi.
[129] Klaus J Hopt (n33) 229.
[130] ibid 230.
[131] ibid 231.
[132] The Codetermination Law of 1951; Horst Siebert, *The German Economy: Beyond The Social Market* (Princeton University Press 2005) 328-9; Horst Siebert (n4) 278; Harald Baum (n10) 15.
[133] The Co-Determination Act 1976; Horst Siebert (n4); Harald Baum (n10).
[134] Klaus J Hopt (n33) 236-7.
[135] Herbert J Spiro, 'Codetermination in Germany' (1954) 48(4) American Political Science Review 1114, 1118-20.

a longstanding tradition of industrial democracy through workers' councils dating back to the Frankfurt Parliament in 1848.[136]

Ideas of the public interest, and need for the state to regulate to protect it, feed into what is often described as the 'stakeholder philosophy' of German company law: the idea that the interest of shareholders alongside those of the enterprise, its workforce and creditors were all various parts of the *aufsichtsrat*'s concern.[137] It is telling that the Stock Corporation Act 1937 provided that the *vorstand* was responsible for not only the shareholders' interest but also those of the workforce and public good.[138] This attitude persists, empirical research shows *vorstand* members still consider themselves bound by a particular duty towards their workforce.[139] These three features: the two-tier board, the stakeholder philosophy and state regulatory involvement remain hallmarks of German company law today. Therefore, the German company has a considerable Kantian inheritance which is reflected in its two-tier board architecture. In providing for labour representation, the notional 'individual' is represented at the highest level of corporate decision-making. This, it is submitted, is a concession to Kantian ethics.

Kant requires us, in the formula of humanity as an end in itself to '[a]ct so that you use humanity, as much in your own person as in the person of every other, always at the same time as an end and never merely as a means'.[140] Therefore, Kant prohibits the individual from being used as an instrument or means to an end where he does not explicitly consent to do so. Conceptualised as an end in himself the individual, or employee, is perceived as the holder of rights and an important party in an economic compact. In other, more familiar words, German enterprise is conceived of as a coalition of capital and labour, the classical factors of production.[141]

3.3 Comparative Analysis

Therefore, German board architecture is built on the tacit assumption that the employee is an end in himself. Whereas English board architecture is built on the tacit assumption that the employee is an instrument of wealth creation to be negotiated with contractually outside the board. The difference in approach is demonstrated by a brief comparison of relevant legislation, company law theory, labour relations and the normative terms of the debate.

---

[136]Horst Siebert (n4) 292-4; Shawn Donnelly, 'The Public Interest and the Company in Germany' in John Parkinson, Gavin Kelly and Andrew Gamble (eds), *The Political Economy of the Company* (Hart Publishing 2001) 90; Markus Roth, 'Employee Participation, Corporate Governance and the Firm: A Transatlantic View Focused on Occuptational Pensions and Co-Determination' (2010) 11 European Business Organization Law Review 51, 62-3.
[137] Klaus J Hopt (n33) 236.
[138] ibid 237; cf Stock Corporation Act 1965.
[139] ibid.
[140] Immanuel Kant (n125) 47.
[141] Manfred Balz (n10) 2.

3.3.1 Statutory Corporate Objectives

The enshrinement of 'enlightened shareholder value' in s 172 Companies Act 2006 is the clearest example of the instrumentalisation of the individual in English corporate law.[142] The section places directors, themselves the agents of shareholders, under a duty 'to promote the success of the company for the benefit of its members as a whole',[143] and therefore establishes a single corporate objective (or end) of shareholder wealth maximization.[144] The section goes on to identify a number of other stakeholders, including but by no means limited to employees, to whom the directors are to have regard in the discharge of their duty in the preceding subsection.[145] Therefore, under the Act the interests of employees are to be considered in so far as they further the success of the company in the long-term interest of its shareholders, and only to that extent: in the words of our thesis: they are the instruments of shareholder value.

The picture in German law is less clear. Whilst the Stock Corporation Act 1937 provided that the management was responsible for the interests of the workforce and the public good alongside that of shareholders, the provision was not retained in the 1965 Act.[146] The reason for the omission was that the *vorsand*'s obligation to the workforce was apparently 'self-evident' under the law.[147] However, the more recent German business judgement rule refers only to the interests of the enterprise.[148] Roth, drawing on reference to looking after 'the best interest of the enterprise' in the German Corporate Governance Code and recent trends in academic literature, argues that German company law now reflects an enlightened shareholder value approach.[149]

It is submitted that the picture is not quite so clear cut. German law distinguishes between 'the company' as a legal form and 'the enterprise' as a whole[150] and the German Corporate Governance Code defines 'the interest of the enterprise' as 'the common duty of the management and supervisory boards to contribute, in accordance with the principles of the social market economy, to the continued existence of the company and to sustainable added value.'[151] Therefore, whilst the boards are under a clear duty to ensure that the company survives and makes a profit this duty is to be exercised within the principles of the social market economy. The ordoliberal notion of a social market economy is 'an economic order attempting to meld the market approach and competition with social protection and equity' with a basic aim of protecting the

[142] HC Deb 6 June 2006, vol 447, col 125.
[143] Companies Act 2006, s 172(1).
[144] Lisa Linklater, 'Promoting success: the Companies Act 2006' (2007) 17 Comp Law 129; Andrew Keay (n8).
[145] Companies Act 2006, s 172(2).
[146] Klaus J Hopt (n33) 237-8.
[147] ibid; Markus Roth (n136) 63; Bruno Kropff, *Aktiengesetz* (Düsseldorf, Verl. Buchh. Des Instituts der Wirtschaftsprüfer 1965) 97ff.
[148] German Stock Corporation Act (AktG), s 93(1)2.
[149] Markus Roth (n136) 64-5.
[150] Christine Windbichler, 'The Public Spirit of the Corporation' (2001) European Business Organisation Law Review 795, 801.
[151] German Corporate Governance Code para 3.8(1).

individual.[152] Therefore, the obligation on the part of management to consider the interests of the workforce alongside those of the shareholders has hardly been exorcised from German corporate law. It is submitted that the conception of the individual as an end in himself remains embedded in the statutory corporate objectives of the German company.

3.3.2 Private verses Public Conceptions of the Company

The long established view of the English company as a private entity, or 'little republic',[153] for the common benefit of shareholders, facilitates the instrumentalisation of the individual within the corporation.[154] Whether in the context of Elizabethan maritime adventures[155] or modern 'nexus of contracts' type business organisation,[156] the company is viewed as a private bargain between shareholders, for their benefit in which no external intervention is warranted.[157] The resilience of this private bargain was demonstrated by the strident rejection of both the Draft Fifth Company Law Directive[158] and the Bullock Committee's proposal[159] which sought to introduce mandatory board-level representation of employees.[160] In contrast, the German view of the company is tempered by considerations of the public interest. The 1937 Act (discussed above) included a general duty on behalf of management to the 'common weal of the people and country'.[161] Public interest concerns lay behind the imposition of the *aufsichtsrat* and codetermination.[162] Codetermination is definitional of the German company, its function, as 'corporatist approach' to governance or 'sub-optimal political compromise', was to engender solidarity and cooperation in the reconstruction of the German economy after both World Wars.[163] Therefore, where the German company progressively came to embrace the public interest, and the interests of its workers, the English company steadfastly remained a private affair. This conceptual distinction underpins the differing approaches to the individual within labour relations.

---

[152] Horst Siebert (n4) 24-5; cf David J Gerber (n23) 232-65.
[153] Susan Watson (n113) 613.
[154] Colin Arthur Cooke (n119); Susan Watson (n113).
[155] Colin Arthur Cook (n119) 55.
[156] Michael C Jensen and William H Meckling, 'Theory of the Firm: Managerial Behaviour, Agency Costs and Ownership Structure' (1976) 3 J Fin Econ 305, 311; Frank H Easterbrook and Daniel R Fischel, 'The Corporate Contract' (1989) 89 Columbia Law Review 1416, 1426.
[157] Shawn Donnelly (n136) 85.
[158] Proposal for a Fifth Company Law Directive founded on article 54(3)(g) of the EEC Treaty concerning the structure of public limited companies and the powers and obligations of their organs, 1972 EC Bull Supp 10 ('1972 Proposed Fifth Directive').
[159] Department for Trade, 'Report of the Committee of Inquiry on Industrial Democracy' (Cmnd 6706, 1977).
[160] Paul L Davies (n123) 377.
[161] Christine Windbichler (n150) 809.
[162] Klaus J Hopt (n33) 230.
[163] Harald Baum (n10) 15; Wanjiru Njoya, 'Employee Ownership and Efficiency: An Evolutionary Perspective' 2004 Industrial Law Journal 211, 235; Klaus J Hopt (n33) 237.

3.3.3 Internalised verses Externalised Labour Relations

Labour relations in England are, and have always been 'external', or outside the board, and often outside the company and initially, the law.[164] It is submitted the external, arm's length nature of the resulting relationship facilitates the instrumentalisation of the individual. Historically, collective bargaining through trade unions was the 'distinctive feature' of the British system, 'the single channel of representation of the interests of the employees' within the company.[165] Davies considers its dominant position causal in the failure of internal forms of representation to take hold.[166] Regardless, the practice of collective labour relations in England has largely taken place between corporate outsiders (trade unions) and company management outside the board. As such, employees have been regarded as a group to be negotiated, and contracted with, at arm's length. The external nature of labour relations was sharpened by the 'Thatcher revolution' of the 1980s where collective bargaining trade union membership waned in favour of individual bargaining.[167] The English worker therefore found himself with a low level of statutory protection, no mandatory internal voice within the company, and—most importantly for our thesis—an individual, contractual relationship at the centre of his employment[168] in place of one largely determined by collective bargaining.[169] In short the employee is, in the words of Easterbrook and Fischel a contractual supplier of labour to be negotiated collectively or individually.[170]

In contrast, by affording board level representation German corporate law internalises the dialogue between labour representatives and the company.[171] Codetermination therefore gives labour interests both control rights over corporate decisions and 'a voice' at the highest level of corporate decision-making.[172] This leads to what Baum refers to as 'a kind of negotiated management... and willingness for long term commitment'.[173] In other words, codetermination fulfils a 'consensus building function between capital and labour'.[174] This long-term commitment is underlined by the wide spread retention of a considerable amount of employee pension contributions within firms as legal capital on the part of workers and higher firm-wide investment in employee training and welfare rights on behalf of the employer.[175] Codetermination internalises the dialogue with the trade unions also: the 1976 Act provides them with

---

[164] Simon Deakin and Gillian S Morris, *Labour Law* (4th edn, Hart Publishing 2005) paras 1.5-9; Paul L Davies (n123) 374-5.
[165] Paul L Davies and Claire Kilpatrick, 'UK Worker Representation After Single Channel' 2004 Industrial Law Journal 121-2; Paul L Davies (n123) 374.
[166] ibid 376-377.
[167] ibid 379.
[168] Frank H Easterbrook and Daniel R Fischel (n156) 1425ff.
[169] Wanjiru Njoya, (n163), 233-5; Paul L Davies (n123) 380-1.
[170] Frank H Easterbrook and Daniel R Fischel (n156) 1429; cf Lord Wedderburn, 'Employees, Partnership and Company Law' 2002 Industrial Law Journal 99, 103, 110-1.
[171] Harald Baum (n10) 15.
[172] ibid 16.
[173] ibid.
[174] Klaus J Hopt (n33) 248.
[175] Markus Roth (n136) 67; Thomas Clarke and Richard Bostock (n43) 246; Shawn Donnelly (n136) 93.

two of the six *aufsichtsrat* seats allocated to labour.[176] Codetermination therefore is an important source of union power, increasing their political leverage and providing positions for their leaders.[177] Therefore, by affording him representation at board level, the notional worker acquires control rights in the German company; rather than carrying out the decisions of others for their profit, the worker is a notional party to corporate decisions and is therefore recognised as an end in himself.

3.3.4 Language of the Discourse

The instrumentalisation of the individual in the English company is unintentionally summarised by Davies when he describes the tendency 'not to view labour law issues in terms of rights'.[178] In company law, great emphasis is placed on the rights of shareholders, a necessary corollary to the duties imposed on directors.[179] It is submitted that this apparent asymmetry of rights is a reflection of the instrumentalisation of the individual: he is an instrument of profit (collectively or individually) and as such is to be bargained with for the corporate gain. In contrast, shareholder value is the end goal, protected by enforceable rights. The extensive set of employment rights absent in English law would be incompatible with the instrumental role of the employee within the corporate context, as if such rights were given effect they would, necessarily, detract from the objective of shareholder value. [180] We find the exact reverse in German law which favours a more rights-based view, from the 'dignity of man' blessed with 'inviolable and inalienable human rights' found in Article 1 of the German Constitution, the *Grundgesetz*[181]to representation rights under codetermination, greater protection of the individual through state welfare provision, a more liberal approach to trade union activity and greater protection of the individual through employment regulation.[182] Therefore, the employee is a holder of rights, an end in himself.

3.4 Conclusion

It has been demonstrated that German and English board architectures rest upon tacit assumptions about the treatment of the individual derived from historically significant philosophies. Such ideas, alongside powerful historical forces, have produced institutions which differ not merely in their legal structures but also in the ideas which underpin them. The insight that the English board architecture reflects the instrumentalisation of the individual in English corporate law and that the German architecture reflects the Kantian view of the individual as an end in himself and therefore a holder of rights is a powerful exegetical tool.

---

[176] Horst Siebert (n4) 287.
[177] ibid 289.
[178] Harald Baum (n10) 378.
[179] Companies Act 2006, s 170-8; Company Directors Disqualification Act 1986; Criminal Justice Act 1993: ss 52-3.
[180] ibid 380-1.
[181] Horst Siebert (n4) 25-6.
[182] ibid 35-7.

It is submitted that one of the reasons behind the passion and colour of the board architecture debate is the difference in starting points or tacit assumptions of the participants. If we conceptualise the proponents of two-tier board architecture as holding, to some extent, the Kantian view and those who favour the unitary board as influenced by utilitarianism, the debate ceases to be a conflict between legal institutions and policy options but a conflict of ideas. Caution, however, must be exercised in elevating the debate to this level, lest it collapses into the equally intractable debate between utililitarians and Kantians. Therefore in the practical terms of the convergence debate, our question becomes, to what extent should the individual be both an end in himself and the means to shareholder profit within corporate governance? The English and German systems, as demonstrated above, differ in their responses.

4 CORPORATE OBJECTIVES

English and German board architectures rest on different tacit assumptions about the corporate objective, or purpose of the firm. This section will articulate these assumptions and demonstrate that the board architectures, and their differences, are better understood in this context. It will be argued that the ideology of shareholder-value pervades both the English company and the unitary board. In contrast, it will be argued that the German corporate objective is multifaceted and serves to promote at least two distinct ends: the generation of wealth for shareholders, on the one hand and the maximisation of employee welfare on the other.

4.1 Shareholder-value and the Unitary Board

Discussion of both the corporate objective and convergence is dominated by Hansmann and Kraakman's 'end of history' hypothesis which asserts the dominance of a shareholder-centred ideology of corporate law.[183] They assert that '[t]here is no longer any serious competitor' to the view that corporate law should seek to increase long-term shareholder value'.[184] Shareholder value is the dominant theory in Anglo-American jurisdictions,[185] it requires simply that the directors 'manage the company in such a way as to ensure the wealth of shareholders is maximised to the full'.[186] Bainbridge divides the theory into two limbs: firstly, the objective of the company should be to maximise shareholder wealth and secondly, shareholders, as the bearers of residual risk, should have ultimate control of the company.[187] According to this theory, shareholder wealth maximisation is the primary imperative of the company and the main responsibility of directors, other responsibilities are 'secondary or derivative' to this overriding pursuit.[188] The interests of other corporate constituencies, such as

[183] Henry Hansmann and Reinier Kraakman (n 62) 33.
[184] ibid 33; cf Antoine Rebérioux, 'European Style of Corporate Governance at the Crossroads: The Role of Worker Involvement' (2002) 40(1) JCMS 111.
[185] Andrew Gamble and Gavin Kelly, 'Shareholder Value and the Stakeholder Debate in the UK' (2001) 9 Corporate Governance: An International Review 110, 110; Andrew Keay (n8) 1.
[186] Andrew Keay (n8) 5.
[187] Stephen Bainbridge, 'Director Primacy: The Means and Ends of Corporate Governance' (2003) 97 Northerwestern University Law Review 547, 573; Andrew Keay (n8) 5.
[188] Andrew Keay (n8) 6; Milton Freidman, 'The Social Responsibility of Business is to Increase its Profits' *The New York Times Magazine* (13 September 1970).

employees, creditors, suppliers and customers are to be protected by contractual and regulatory means rather than corporate governance.[189]

The historical development of the English company as a private undertaking for the benefit of its members is discussed previously.[190] However, the link between shareholder value and the unitary board does not require historical explanation: in law the directors of an English company are appointed by, or on behalf, of shareholders and are under a range of enforceable statutory duties to act in the interests of shareholders.[191] Shareholders are the only group to whom UK company law and corporate governance afford board-level representation, control rights and legal protections.[192] The agency problems arising from the separation of ownership and control dominate corporate governance scholarship; such a focus presupposes the corporate objective is the maximisation of shareholder value.[193]

The most recent formulation of shareholder value is the 'enlightened shareholder value' (ESV) concept enshrined in s 172 Companies Act 2006. The section imposes a general duty upon directors 'to promote the success of the company for the benefit of its [shareholders] as a whole'.[194] The 'enlightened' aspect follows: it identifies a number of other corporate constituents, including but by no means limited to employees, to whom the directors are to have regard in the discharge of their duty in the preceding subsection.[195] It is often argued that ESV is a compromise between shareholder-value and the interests of the other corporate constituents, promotes accountability and uses market forces to move towards greater social responsibility.[196] This view is erroneous, s 172 clearly identifies shareholders as the master constituency, the other constituents are merely to be considered and used as means in the long-term service of the shareholders.[197] Therefore, shareholder value remains the corporate objective in English law and the 'end of history' hypothesis is correct to this extent.[198]

---

[189] Henry Hansmann and Reinier Kraakman (n62) 35.
[190] Colin Arthur Cooke (n119) 55.
[191] Companies Act 2006, s 170-8; Company Directors Disqualification Act 1986; Criminal Justice Act 1993: ss 52-3.
[192] ibid.
[193] John Armour, Henry Hansmann and Reinier Kraakman, 'Agency Problems and Legal Strategies' in Strategies' in Reinier Kraakman, John Armour, Paul Davies, Luca Enriques, Henry Hansmann, Gerard Hertig, Klaus Hopt, Hideki Kanda and Edward Rock, *The Anatomy of Corporate Law: A Comparative and Functional* Approach (2nd edn, OUP 2009) 35-7; Klaus J Hopt and Patrick C Leyens, 'Board Models in Europe: Recent Developments of Internal Governance Structures in Germany, the United Kingdom, France and Italy' in Levinus Timmerman (ed), *The Verenigde Oostindische Compagnie 1602 - 400 Years of Company Law* (Kluwer 2004) 281, 283.
[194] Companies Act 2006, s 172(1).
[195] Companies Act 2006, s 172(2).
[196] Cynthia A Williams and John M Conley, 'An Emerging Third Way?: The Erosion of the Anglo-American Shareholder Value Construct'(2005) 38 Cornell International L J 493.
[197] Lorraine E Talbot, 'A Contextual Analysis of the Demise of the Doctrine of *ultra vires* in English Company Law and the Rhetoric and Reality of Enlightened Shareholders'2009 Comp Law 323.
[198] Lisa Linklater, 'Promoting success: the Companies Act 2006' (2007) 17 Comp Law 129; Andrew Keay (n8); Henry Hansmann and Reinier Kraakman (n62); cf Michael C Jensen, 'Value Maximisation, Stakeholder Theory, and the Corporate Objective Function' (2001) 14(3) Journal of Applied Corporate Finance 8.

4.2 Stakeholder Theory, Codetermination and the Persistence of Labour

Stakeholder theory is rooted in the ideal that 'all parties work together for a common goal and obtain shared benefits' and therefore the corporate objective (and the duty of managers) is to create optimal value for all stakeholders; stakeholders are social actors who might be regarded as parties who can affect or are affected by a company's decisions.[199] Stakeholders include Hansmann and Kraakman's corporate constituents[200] but also wider interests such as those of the government, local communities, state and even 'the environment'.[201] Stakeholder theory does not give any guidance on how to prioritise between stakeholders when their interests conflict.[202]

The German two-tier board is commonly held up as an institutional example of stakeholder theory in action on account of the presence of both labour and shareholder representatives on the *aufsichtsrat*.[203] It is submitted that the architecture only reflects stakeholder interests to the extent that it represents labour and capital interests directly, bank employees and business partners will often sit as shareholder representatives and the wider role of the *aufsichtsrat* in networking is not to be underestimated.[204] However, to argue codetermination is a vindication of stakeholder theory is inaccurate: it privileges shareholders and labour by affording them control rights which could easily be extended to other interest groups.

Hansman and Kraakman convincingly argue that the German system is actually 'labour orientated' by virtue of the codetermination settlement, collective bargaining and tradition of industrial democracy.[205] They argue that stakeholder models, behind the rhetoric, are in reality a formulation of the labour-orientated model merely extended to other stakeholders.[206] This argument is descriptively convincing owing to the privileged position of German employees. Hansman and Kraakman further argue that the labour model has failed as a model for convergence, citing the EU retreat from the Fifth Directive and therefore lacks normative appeal.[207] They submit that although employee involvement in decision-making may mitigate some inefficiencies of labour contracting, these gains will not exceed the costs of the inefficiencies the system entails.[208] These two conclusions are correct. However, as they concede, neither argument invalidates the practice of codetermination in Germany, or amounts to a serious challenge to its importance in relation to the corporate objective.[209] The

[199] Andrew Keay (n8) 6-7; cf Gavin Kelly and John Parkinson (n4) 130.
[200] Henry Hansmann and Reinier Kraakman (n62) 35.
[201] Bob Tricker (n1) 229-31.
[202] Andrew Keay (n8) 8.
[203] Thomas Clarke and Richard Bostock (n43) 231, 245; Klaus J Hopt and Patrick C Leyens (n193) 288; Florian Schwarz (n4) 128-130.
[204] Klaus J Hopt and Patrick C Leyens (n193).
[205] Henry Hansmann and Reinier Kraakman (n62) 37-8.
[206] ibid 42.
[207] ibid 38-9.
[208] ibid; cf Klaus J Hopt and Patrick C Leyens (n193) 290-1; Klaus J Hopt (n33) 236-8.
[209] Henry Hansmann and Reinier Kraakman (n62) 39.

persistence of labour within the corporate objective is reflected in German law through the ordoliberal notion of the social market and was demonstrated in the previous section. It is therefore submitted that the German corporate objective is split between shareholder value and employee welfare.

4.3 Comparative Analysis

It is submitted that the difference in corporate objective provides a compelling explanation for not merely the advantages and disadvantages discussed in section two, but also the different priorities and critical weighting respective authors attach to the various features. Proponents of the unitary board stress efficiency and maximisation of shareholder-value.[210] In contrast, advocates of the two-tier system attach greater weight to achieving employee welfare through corporate governance mechanisms, at the expense of shareholders. Studies indicate its costs may be as high as 20% of shareholder-value.[211]

4.3.1 Decision-making

Proponents of the unitary board underline its flexibility and responsiveness. It is able to meet frequently, in full or sub-committee, institute change quickly and take painful measures such as restructuring; however, such boards can tend towards short-term decision-making.[212] In contrast, two-tier board are less able to make quick decisions as they meet less frequently and contain the entrenched interests of existing employees who will often be change averse for reasons of employment security.[213] In other words, perhaps the codetermined board has 'the engine of a lawnmower and the breaks of a [BMW]'.[214]

The fast decision-making in a unitary board allows it to pursue shareholder-value, responding quickly to market changes and effecting large changes within the business, such as restructuring, redundancies and plant closures with unity, precision and speed. The constraints in such matters are external in the form of regulation and contractual obligations. In contrast, the codetermined board, often portrayed as weaker,[215] is unable to pursue shareholder value with the same vigour and instead must engage in 'sub optimal compromises'[216] and 'corporatist' decision-making by appeasing the labour representatives.[217] Therefore, its decision-making process is more drawn out and negotiated and it is less able to react quickly to the market conditions and will seek to avoid painful measures such as redundancies, restructuring and plant closures are

[210] Mark J Roe, 'Modern Politics and Ownership Separation' in Jeffrey N Gordon and Mark J Roe (eds), *Convergence and Persistence in Corporate* Governance (CUP 2004) 252, 259.
[211] ibid; Felix R FitzRoy and Kornelius Kraft, 'Economic Effects of Codetermination' (1993) 95 Scand J Econ 365.
[212] Horst Siebert (n4) 290; Thomas Clarke and Richard Bostock (n43) 244.
[213] Horst Siebert (n4) 290.
[214] Jonathan Lynn and Antony Jay (eds), *The Complete Yes Prime Minister: The Diaries of the Right Hon. James Hacker* (BBC Books 1989) 140.
[215] Mark J Roe (n210) 258.
[216] Wanjiru Njoya (163) 233.
[217] Harald Baum (n10) 1, 15.

commonly avoided due to the inherently conservative interests of the labour members. The result is management is forced into taking decisions to promote the employee interest and the long-term management of the business, at the expense of short-term profit. From the perspective of shareholder-value, where the corporate objective subordinates the interest of the shareholder to the employee the features of the German system are genuine disadvantages. However, when viewed from the perspective of a bifocated corporate objective, the criticisms lose critical force.

4.3.2 Labour Relations and Human Capital

This difference in corporate objectives and decision-making mechanisms offers a compelling explanation of the difference in labour relations and approach to human capital between English and German companies. Codetermination is said to yield better labour relations, and less industrial action, as employees have 'a voice' within the board.[218] In reality, the presence of labour representatives allows employees to affect corporate policy, and reallocate corporate risk, in their favour.[219] Therefore, labour is able to engage more effectively in rent seeking behaviour and will, as a result be more satisfied than in an Anglo-Saxon firm which seeks to direct wealth exclusively to shareholders.[220] Therefore, the more harmonious labour relations found in codetermined companies are not so much as a result of abstract representation but because the corporation is institutionally more inclined to act in their interest. It is submitted that the more harmonious labour relations, while advantageous to a German company, are bought at a high cost to shareholders, and is only justifiable if the purpose of the company includes the welfare of its employees.

The recognition of employee's intelligence, and skill as principal assets of the business and the correspondingly higher investment in human capital, on the part of German firms, and the employees themselves, is, it is submitted, a result of the bifocated corporate objective.[221] Investment in human capital can be generic (fully redeployable to alternative uses without loss of value) or firm-specific (where they are not redeployable).[222] In the former case where the company finances employee development it is exposed to the risk of opportunistic behaviour on the part of the employee and in the latter the risk of opportunism is borne by the employee.[223] The control rights afforded by codetermination protect German workers against opportunism in the latter case. Under the English system, workers in the latter position would be expected to protect themselves through contractual provisions such as demanding a higher price for their labour or safeguards such as insurance.[224]

[218] Thomas Clarke and Richard Bostock (n43) 245.
[219] Horst Siebert (n4) 289.
[220] Mark J Roe (n210) 257-60.
[221] Thomas Clarke and Richard Bostock (n43) 246.
[222] Gavin Kelly and John Parkinson (n4) 119.
[223] ibid.
[224] ibid 120.

4.3.3 Control of Discretionary Powers and Conflicts of Interest

Mäntysaari argues that the English unitary board has an 'excess of powers' as a result UK company law's failure to separate management and control functions institutionally.[225] In contrast, he asserts the German system provides for more effective supervision and control over management through institutional separation and stresses that this control is vital for the protection of shareholders.[226] Hopt submits the opposite, that codetermination leads to fractionalization evidenced by the practice of separate labour and shareholder pre-meetings before the board meeting and reluctance for shareholders to openly criticise management.[227]

Similarly, Hopt and Baums and Scott point out that German law does not adequately regulate for conflicts of interest between controlling shareholders, directors and the corporation.[228] Labour representatives, and particularly those who are also union representatives, face considerable conflict of interest when calling for a strike against the company upon whose *aufsichtsrat* they sit.[229]

The need for greater supervision of the *vorstand* is explained by reference to the corporate objective and agency theory. As Hopt states, the 'widening the responsibility of the board beyond shareholders not only acerbates the agency problem for the latter but adds the agency problem of labour and gives an excuse for respecting neither of these responsibilities'.[230] Therefore, the German system requires more constraint on discretionary powers, and closer supervision, because of the agency problems created by the bifocation of the corporate objective and codetermination. In contrast, the unitary board in pursuit of shareholder value only creates the singular agency problem of shareholder-director and does require redress through an internal balancing mechanism. Instead, it may be addressed through the general meeting, disclosure requirements, statutory remedies and the disciplines of the markets for securities and corporate control. The extent to which these arrangements succeed is, thankfully, beyond our argument.

4.4 Conclusion

It has been demonstrated that English and German boards of directors are moved by different corporate objectives. If to forget one's purpose is the commonest form of stupidity, then participants in the board architecture debate must be careful to consider their tacit assumptions about the purpose of the institutions in question.[231] It has been established that the advantages and disadvantages lose much of their critical bite when assessed against the differing corporate objectives, as features which promote

[225] Petri Mäntysaari (n4) 401.
[226] ibid 403-4.
[227] Klaus J Hopt (n33) 247.
[228] Theodor Baums and Kenneth E Scott (n74) 19; Klaus J Hopt (n33) 238.
[229]Jean J Du Plessis and Otto Sandrock, 'The rise and fall of supervisory codetermination in Germany?' 2005 ICCLR 67, 75.
[230] Klaus J Hopt (n33) 237.
[231] Friedrich Nietzsche (n28) para 206.

shareholder value at the expense of employee welfare, will be more acceptable in Anglo-Saxon circles than German. Similarly, features which promote employee welfare at the expense of shareholder value may be considered inefficient and find themselves the target of criticism in Anglo-Saxon circles.[232]

With the application of the corporate objective to the debate, it ceases to be between the various features of the board architectures, but comes to the more fundamental question: in whose interest should the company be run? The English and German systems, as demonstrated above, differ in their answers. In the practical terms of the convergence debate, it once again becomes a question of accommodation. The direction of scholarship, in both Germany and England, as well as elsewhere, favours a corporate law organised one or another variations of the shareholder-value imperative.[233] Institutional, as well as ideological, convergence will be possible only to the extent that German law and institutions can adapt to a conception of the company orientated around this objective. However, the use of corporate governance to achieve social market aims remains a key aspect of German law.[234] Accommodation could be achieved through the use of labour law, and other non-company law regulatory means, to maintain the social market and allow for corporate convergence.[235] However, there is little appetite for such far reaching reforms.[236]

5 MARKET IDEOLOGY

It has been argued that English and German board architectures rest on different tacit assumptions about the individual and corporate objective. This section seeks to provide wider context by arguing that the respective architectures are manifestations of different market ideologies. That there is no alternative to liberalism is a truth universally acknowledged; however, it is argued that the dominant forms of liberalism underpinning the English and German economies are different and these differences enlighten our understanding of the differences between the architectures.[237] It is not argued that English and German economy is a perfect, and exclusive, product of any single ideology.[238] Rather it is argued that these ideas are, and have been, prominent in the English or German intellectual *zeitgeister* and consequently influential in the development of our economic institutions and our understandings thereof.[239] Therefore, the analysis of the board architecture debate in this context may prove enlightening.

---

[232] Michael C Jensen (n198).
[233] eg Henry Hansmann and Reinier Kraakman (n62); Theodor Baums and Kenneth E Scott (n74); Markus Roth (n136) ; Horst Siebert (n4).
[234] Klaus J Hopt (n33) 238.
[235] Lord Wedderburn (n170) 110-1; Michel Aglietta and Antoine Rebérioux (n3) 54-7.
[236] Klaus J Hopt (n33) 255.
[237] Raymond Geuss, 'Liberalism and Its Discontents' in Raymond Geuss, *Outside Ethics* (Princeton University Press 2005) 11; John Dunn, *Western Political Theory in the Face of the Future* (2nd ed, CUP 1993) 29-30; Stephanie Lee Mudge, 'The State of the Art: What is Neo-liberalism?' (2008) 6(4) Socio-Economic Review 703, 715-18.
[238] Friedrich August von Hayek, *The Road to Serfdom* (Routledge 2001) 4-5; Stephanie Lee Mudge, (n1) 711-2.
[239] David J Gerber (n23) ix, x, xxi.

5.1 Anglo-Saxon Neoliberalism

A prominent proponent of Anglo-Saxon neoliberalism is FA Hayek, 'who assumed the status of a divine in the Thatcherite scriptures'.[240] So much so that, on her only visit to the Conservative Research Department Thatcher produced a copy of Hayek's *The Constitution of Liberty* and announced, 'This is what we believe'.[241] Hayek saw himself as a defender of the classical liberalism of nineteenth century England, or 'Whiggism'.[242] Therefore his work draws heavily on the older liberal tradition of Adam Smith,[243] William Gladstone, JS Mill, Lord Acton and others.[244] For Hayek, the ultimate political virtue was a negative form of personal liberty,[245] typified by a minimisation of coercion or interference.[246] He that argued such personal and political freedom could only be sustained through economic freedom.[247] Realising economic freedom mandated a free competitive market[248] supported by a rule of law based legal system[249] designed to preserve competition and make it operate as beneficially as possible.[250] Therefore, government was to minimise its market interventions.[251] However, Hayek accepted that it had a role in preserving freedom of contract, property rights, protecting against fraud and providing social services which could not be policed by competition.[252] Consequently, he advocated a strong but limited state[253] and opposed central planning, not merely on grounds of efficiency[254] but because it expanded the role of government at the expense of personal and economic freedom.[255] 'The chief evil is unlimited government, and nobody is qualified to wield unlimited power.'[256]

Hayek's influence on Thatcher is clear in her statement that the purpose of government was to provide a 'framework of stability' in constitutional, legal and economic terms, within which families and businesses could flourish.[257] This project included liberating the market from the monopolies of the state, labour and business as well as shrinking

[240] Stephen Evans, 'Thatcher and The Victorians: A Suitable Case for Comparison?' (2002) 82 History 601, 606; Kieron O'Hara, 'The Iron Man: Is Cameron True Blue?' (2007) 14(3) Public Policy Research 181, 181; GR Steele, 'There Is No Such Thing As Society' (2009) 29(4) Economic Affairs 85, 85-6.
[241] John Ranelagh, *Thatcher's People: An Insider's Account of the Politics, the Power, and the Personalities* (Fontana 1992) ix; cf Stephanie Lee Mudge (n237) 718.
[242] Friedrich August von Hayek (n42) (Routledge 2006) 352-3.
[243] cf Friedrich August von Hayek, 'The Use of Knowledge in Society' (1945) 35(4) The American Economic Review 519.
[244] Friedrich August von Hayek (n42) 52-3.
[245] ibid 11-12.
[246] ibid 18-19.
[247] Friedrich August von Hayek (n238) 13.
[248] ibid 37-8.
[249] Friedrich August von Hayek (n42) 193.
[250] Friedrich August von Hayek (n238) 39.
[251] Friedrich August von Hayek (n42) 193.
[252] Friedrich August von Hayek (n238) 38-41, 215; Friedrich August von Hayek (n42) 248ff.
[253] Friedrich August von Hayek (n42) 348.
[254] Friedrich August von Hayek (n243) 521ff.
[255] Friedrich August von Hayek (n42) 348.
[256] ibid 348.
[257] Stephen Evans (n240) 605.

the state, minimizing public expenditure and limiting intervention in the market.[258] In the words of Edward Heath, the Conservative Party had 'adopted Manchester Liberalism of about 1860'.[259] Thatcher's revolution irreversibly changed the nature of the British economy and this neoliberal ideology has pervaded the policies of successive governments.[260]

Therefore, the intellectual climate in which the English company and corporate governance emerged was deeply influenced by individualistic, liberal ideas. The nexus of contracts notion of the company as a private entity, participating in a free competitive market by means of freedom of contract is embodied in its unitary board.[261] This approach necessitates both shareholder value objective and the exclusion of non-shareholder representatives from the board as a vindication of property rights.[262]

5.2 German Ordoliberalism

German neoliberalism, otherwise known as ordoliberalism, developed from the circle of legal and economic scholars led by Walter Eucken in Freiberg from 1933.[263] Following the war the ordoliberals, and their ideas, came to shape the social market economy (SME).[264] Ordoliberals, such as Ludwig Erhard, later Minister for Economics and Chancellor, occupied senior administrative positions during the allied occupation and ensured the German economy was rebuilt in line with ordoliberal ideas.[265] Although the ordoliberals and Freiberg had strong personal links with Hayek, their thought is distinct.[266] As for Hayek, ordoliberalism works from the basic starting point that competition is necessary for economic wellbeing and that economic freedom is a prerequisite for political freedom.[267] However, whereas Anglo-Saxon liberalism is rooted in economic notions, ordoliberalism is rather to be viewed as a 'humanistically-based intellectual orientation'.[268] 'The desideratum for a socially equitable order was inextricably weaved into the quest for the best economic order of a society with its ever changing technologies and human expectations.'[269] The key features of the ordoliberal

[258] ibid 606; Friedrich August von Hayek (n238) 204-6; Stephanie Lee Mudge (n237) 718; Mathieu Hilgers, 'The Historicity of the Neoliberal State' (2012) 20(1) Social Anthropology 80, 80-1.
[259] HL Deb 23 January 1985, vol 459, cols 252-3.
[260] John Gray, 'Blair's Project in Retrospect' (2004) 80(1) International Affairs 39; Stephanie Lee Mudge (n237) 721.
[261] Frank H Easterbrook and Daniel R Fischel (n156) 1426; Friedrich August von Hayek (n42) 241.
[262] Friedrich August von Hayek (n42) 109-10, 122-4, 241.
[263] David J Gerber (n42) 29-31; Victor J Vanberg (n42).
[264] ibid 58-9.
[265] ibid 60-1; Konrad Zweig, *The Origins of the German Social Market Economy: The Leading Ideas and Their Intellectual Roots* (Adam Smith Institute 1980) 15-18; Michael A Peters, 'Education, Power and Freedom: Third Way Governmentality, Citizen-Consumers and the Social Market' (2010) 2(1) Contemporary Readings in Law and Social Justice 15, 23.
[266] Viktor J Vanberg, 'Hayek in Freiburg' (2001) Freiburg Discussionpapers on Constitutional Economics 12/1; David J Gerber (n42) 32.
[267] David J Gerber (n42) 36; Werner Bonefeld, 'Freedom and the Strong State: On German Ordoliberalism, (2012) 17 New Political Economy (forthcoming) available at <http://www.tandfonline.com/doi/abs/10.1080/13563467.2012.656082> accessed 20 April 2012 1-2.
[268] Reinhard Behlke, *Der Neoliberalismus und die Gestaltung der Wirtschaftsverfassung in der Bundersrepublik Deutschland* (Duncker & Humbolt 1961) 38.
[269] Konrad Zweig (n265) 8.

SME were a free competitive economy, a strong state, ingrained entrepreneurship, private property, freedom of contract and the free price mechanism intended 'to prevent the proletarianisation of social structures'.[270] Ordoliberal principles were reaffirmed by the German government in 1979[271] and they remain prominent in the German and European *zeitgeister*, not least by virtue of their influence in competition law.[272] The continuing influence of ordoliberalism is evidenced by the approach of Angela Merkel's administration to the present Eurozone crisis.[273] Therefore, the contemporary German company developed in an intellectual climate, which was and remains ideologically distinct from that of its English counterpart.

It would be a careless sleight of hand to attribute board-level codetermination directly to the ordoliberal project, as its intellectual antecedence is rooted in social democratic ideology.[274] Both Eucken and Böhm rejected codetermination as a collusion between workers, unions and entrepreneurs at the cost of the consumer, and an obstruction of the market mechanism.[275] However, it is argued that codetermination has been assimilated into the SME ideological compact and that its survival is best understood in the intellectual context of the SME.

5.3 Comparative Analysis

The differences between neoliberal and ordoliberal ideas provide a convincing rationale for the differences in board architecture; further analysis of the distinctive features of the SME in this context will also demonstrate SME-based justifications for codetermination.

5.3.1 Competition and the State

Competition is of central importance to both Hayek and the ordoliberals in creating and maintaining a free market and economic and personal liberty. The ordoliberals believed that individual greed, 'though required to oil the wheels of competition, is all consuming to the extent that it destroys its own foundation'.[276] Preventing this was a political task to be assigned to a constitutional economic order under the protection of a strong state.[277] Therefore the ordoliberal state was charged with creating a stable

[270] Werner Bonefeld (n267) 2.
[271] Konrad Zweig (n265) 33.
[272] David J Gerber (n42) 60ff; David J Gerber (n23); Nicola Giocoli, 'Competition Verses Property Rights: American Antitrust Law, The Freiburg School, and the Early Years of European Competition Policy' (2009) 5(4) Journal of Competition Law and Economics 747, 772.
[273] Hans Kundnani, 'The eurozone will pay a high price for Germany's economic narcissism' *The Guardian* (London, 6 January 2012) <http://www.guardian.co.uk/commentisfree/2012/jan/06/eurozone-germany-ordoliberalism> accessed 23 August 2012; Antoine Vauchez, 'The economic order that inspires Merkel' *Libération* (Paris, 6 December 2011) < http://www.presseurop.eu/en/content/article/1262991-economic-order-inspires-merkel> accessed 23 August 2012.
[274] Abraham Shuchman, 'Economic Rationale of Codetermination' (1956) 10 Industrial and Labour Relations Review 270.
[275] Konrad Zweig (n265) 31.
[276] Werner Bonefeld (n267) 2.
[277] ibid 2; Victor J Vanberg (n42) 3, 5.

framework of rules and institutions within which the market would operate as well as 'seeking to improve the resulting economic order in an *indirect* manner, by reforming the rules of the game'.[278] This necessitated a strong and well enforced competition law to prevent the degeneration of competition and maintain the conditions under which it would flourish.[279] Competition law was directed at private economic power, which if left unchecked threatened not merely competition, but the independence of the state.[280] Hayek argued that there was no need for the state to play a major role in maintaining the conditions of competition.[281] Instead, Hayek limited himself to the argument for the state as the locus of the rule of law and the legal framework for market exchange relations.[282]

Therefore, the ordoliberal state is charged with the task of controlling private economic power for the common good, but is prevented from doing so by *direct* intervention. It is submitted that codetermination also services this political goal. By allocating a proportion of control-rights in large companies to labour representatives, codetermination *indirectly* constrains the discretion of entrepreneurs and capitalists in the conduct of their business.[283] Therefore, codetermination is part of this ordoliberal 'balanced institutional design'.[284] It can be justified as a further aspect of the economic constitutional framework designed to limit the pursuit 'greed' and 'selfishness' of private power in the interests of all.[285] In contrast, such controls are not ideologically necessary for Hayek who considered economic freedom the essence of liberalism.[286] Therefore, it is unsurprising that England which is more ideologically tolerant of private power and less concerned with maintaining the conditions of competition did not develop such structures to constrain the pursuit of shareholder-value.

5.3.2 Social Policy

Ordoliberal and Hayekian approaches to social policy are vastly different. While Hayek accepted the need for a social minimum in addition to the free market, the ordoliberals argued the 'safety of market liberty presupposes the strong state as the provider of... social and ethical frameworks to embed entrepreneurialism... into society at large'.[287] Müller-Armack and Erhard's formulation of the SME places increased emphasis on a 'market-confirming'[288] social policy in the ordoliberal project.[289] Their social policy is intended to 'undercut demands for collective forms of welfare provision in favour of a

[278] Victor J Vanberg (n42) 7.
[279] David J Gerber (n42) 50-51.
[280] Werner Bonefeld (n267) 19.
[281] David J Gerber (n42) 32; David J Gerber (n23).
[282] Werner Bonefeld (n267) 3.
[283] Herbert J Spiro (n135).
[284] Ulrich van Suntum, Tobias Nöhm, Jens Oelgemöller and Cordelius Ilgmann, 'Walter Eucken's Principles of Economic Policy Today' 2011 CAWM Discussion Paper No 49, 17.
[285] Nicola Giocoli (n272) 769; Werner Bonefeld (n267) 19.
[286]Friedrich August von Hayek (n42); Werner Bonefeld (n267) 4.
[287] Werner Bonefeld (n267) 3.
[288] ibid 12.
[289] Ulrich van Suntum, Tobias Nöhm, Jens Oelgemöller and Cordelius Ilgmann (n48) 8; David J Gerber (n42) 60.

human economy of self-responsible social enterprise.'[290] They sought not merely to placate the proletariat but 'to do away with them' through a market which increased wealth, stable employment and security of wage income.[291] For Röpke, the market went 'beyond demand and supply', it was about re-rooting workers in conditions of 'self-provisionment and property [that] ... will enable [the nation] to withstand even the severest shocks without panic or distress.'[292] Therefore, the market was a 'social instrument' to be used to promote the acquisition of property by the working class and limit the 'social devitalisation and spiritual abandonment' of the welfare state.[293] In contrast, Hayek, as discussed above, did not share these ideas and saw some form of non-market welfare state as inevitable.[294] Instead, he was critical of the idea of 'social justice' and argued that it contradicted the essence of a market economy and entailed a 'dangerous and seductive enunciation of tyranny'.[295]

Codetermination can therefore be conceptualised as an element of the ordoliberal SME as it seeks to integrate human beings and enterprise. As noted earlier, codetermination has been credited with achieving to some extent all of the goals listed above, and indeed as Roe notes up to 20% of the wealth that would, in an Anglo-Saxon firm, otherwise go to shareholders is directed to labour as a result.[296] Such a direction of wealth furthers the SME aim of embourgeoisement through an indirect market mechanism rather than wealth redistribution or control of labour supply. In contrast, the absence of labour representation in the unitary board reflects, in part, the dichotomy between the market and social policy in both classical and Hayekian liberalism.[297] While Hayek tolerates social policy, for him it must either be subsumed into the market in entirety or completely separate lest society advances down the road to serfdom.[298] Therefore, the board, as the directing component of a market actor must be remain free to compete in the market and pursue shareholder-value.

5.4 Conclusion

It has been demonstrated that the English and German board architectures are underpinned by different forms of liberalism. Further, it has been demonstrated that structural differences are explicable by reference to divergent ideological notions of the role of the state in the market, the nature of competition in the market and the relationship between social policy and the market.

In the terms of the convergence debate, the positive contribution of this section is to show that the differences between the respective ideologies are not of a categorical nature in the way that all liberals, properly so called, differ in their convictions from

[290] Werner Bonefeld (n267) 2.
[291] ibid 5.
[292] ibid 12.
[293] ibid 5, 8.
[294] Friedrich August von Hayek (n238) 38-41, 215; Friedrich August von Hayek (n42) 248ff.
[295] Werner Bonefeld (n267) 13; Friedrich August von Hayek (n42) 228.
[296] Mark J Roe (210) 252, 259.
[297] Werner Bonefeld (n267) 11.
[298] Friedrich August von Hayek (n42) 224-5, 248-9; Friedrich August von Hayek (n238).

socialists.[299] Rather, the differences lie in placing of institutions and policies in relation to the market and the necessary prerequisites to protect personal freedom. If Vanberg's reconciliation of Hayekian evolution and spontaneous order with ordoliberal economic constitutionalism is an indication of future developments, these ideologies may prove to differ merely by degree and emphasis, therefore a degree of future theoretical accommodation may precipitate practical convergence.[300]

6 CONCLUSION

The intellectual genealogy ventured in the preceding sections is necessarily incomplete in both breadth and depth owing to the nature of the enterprise and constraints of space. However, the preceding analysis establishes two distinct and cohesive intellectual traditions and their role in shaping English and German board architectures and the contemporary debate surrounding them. Mill's utilitarian individualism, shareholder value and Hayekian liberalism form three strands in a distinctly Anglo-Saxon tradition. In contrast, the German tradition is characterised by the Kantian notion of the individual within a community which complements the less individualistically focused ideologies of stakeholder theory and ordoliberalism.[301] Despite the brevity of our exposition, several conclusions arise.

The board architecture debate is inadequate in focusing on economic analysis or the advantages and disadvantages of the respective systems because it ignores the ideas, history and politics behind the two systems. Therefore, the debate owes its intractable nature not to straightforward differences of opinion but to largely unarticulated starting points or tacit assumptions three of which have been demonstrated in this paper. Board architecture is one element of a wider ideological system and political tradition that extends well beyond the conventional boundaries of corporate governance. The resolution of this debate requires more than a mere policy choice or string of technical regulations. The intractable debate is therefore an aspect of a larger European ideological divergence and cannot be truly solved until an ideological consensus is established, not least in relation to the treatment of the individual within the corporation, the proper purpose of a corporation and the ideology and purposes of the market.

Such an ideological consensus is a political issue of fundamental importance that can only be legitimately established by agreement at high levels, and through open, democratic means. Establishing such a consensus would be a prerequisite for establishing a complete European 'economic constitution' within which corporate convergence could emerge.[302] This conclusion has implications for both the existing debate and the direction of European policy.

---

[299] Friedrich August von Hayek (n42) 351-5.

[300] Victor J Vanberg, 'Hayek's Legacy and the Future of Liberal Thought: Rational Liberalism Verses Evolutionary Agnosticism' (1994) 14(2) Cato Journal 179.

[301] Immanuel Kant (n40) 37.

[302] Wolf Sauter, 'The Economic Constitution of the European Union' (1998) 4 Columbia Journal of European Law 27, 46ff; Viktor J Vanberg (n42) 3ff; David J Gerber (n42) 69-72.

6.1 The Existing Debate Is Irrelevant

The existing debate is irrelevant because the answers it provides are addressed to an irrelevant question: which system is preferable? The advantages and disadvantages approach is irrelevant because it lacks an Archimedean standpoint. The strength of either argument depends on the normative position of the audience and, as has been demonstrated, the proponents have vastly differing normative standpoints, or tacit assumptions. Therefore, the only possible solution to this aspect of the debate lies in the adoption of a single normative standpoint. This standpoint could only cohesively be found in the aforementioned, as yet elusive, ideological consensus.

The economic analysis, in contrast, does provide an Archimedean standpoint in terms of efficiency. However, efficiency alone is an unsatisfactory tool with which to resolve the debate because it is incapable of measuring, or even accommodating, all of the public goods that the respective systems yield. Therefore, whilst it provides a solution, the solution is not predicated upon all germane considerations. In short, while it is a coherent tool, it is overly reductive.

In reality, board architecture is intimately tied to deeper and wider political and economic questions, three of which we have considered. Therefore, any attempt along the lines of the original Proposed Fifth Directive to impose the German system into England, or vice versa, is naive because it ignores the intellectual and historical accoutrements of the system.[303]

6.2 Means of Convergence

A straightforward choice between the two systems, resulting in unification under a single practical system, is neither practical nor desirable. Therefore, two avenues to achieve convergence remain: the compromise achieved in the SE and the process of regulatory competition emerging through freedom of establishment rights.[304] The EU's present path of pursuing both is acceptable, however, it is submitted that regulatory competition is the only meaningful way of achieving a functional convergence.

6.2.1 Societas Europaea

If the SE is to be assessed as a means of achieving convergence, in its present form, it is an abject failure.[305] Instead of providing a truly supra-national corporate form anchored in European law, as was the original intention, it is anchored in the law of

[303] Proposal for a Fifth Company Law Directive founded on article 54(3)(g) of the EEC Treaty concerning the structure of public limited companies and the powers and obligations of their organs, 1972 EC Bull Supp 10 ('1972 Proposed Fifth Directive').
[304] Consolidated Version of the Treaty on the Functioning of the European Union [2010] OJ C83/47 ('TFEU'), Arts 49, 54.
[305] Council Regulation 2001/2157/EC of 8 October 2001 on the Statute for a European company (SE) [2001] OJ L 294/1; Council Directive 2001/86/CE 8 October 2001 supplementing the Statute for a European company with regard to the involvement of employees (SE Employee Involvement Directive) [2001] OJ L 294/22.

member states.[306] Therefore, instead of achieving convergence through by-passing national law, it has actually been subsumed into national law.[307] Rather than supplying a single new corporate form to the existing twenty-seven it adds a further twenty-seven.[308]

Within the SE employee representation at board level is determined by an intricate mixture of national and European rules.[309] The operative principle is 'no export' of employee participation and 'no escape'.[310] Clear priority is given to preventing escape, much to the relief of the German labour unions.[311] The upshot, for our purposes, is that a German company seeking to avoid codetermination will be unable to do so via transformation[312] and will only be able to do so in limited circumstances via cross-border merger.[313] Such mergers between British and German companies, where the German company comprises 25% or more of the total post-merger employees, will be mandated to retain the German level of employee participation, even if the seat of the company is located in Britain.[314] Therefore, rather than providing a means through which board structures will converge, the SE stands to exacerbate the problem in practice. Nevertheless the SE is a political compromise, of limited interest to companies[315] and represents some level of mutual accommodation as well as a restatement of the problem in legislation. It therefore represents little hindrance to convergence through regulatory competition and goes some way towards reassuring, and distracting, member states that would otherwise resort to corporate protectionism.

6.2.2 Jurisdictional Competition

In the absence of unification through economic constitutionalisation[316] regulatory competition is not a 'spectre', to be avoided because of an inevitable race to the bottom, but a reality.[317] It is the inevitable byproduct of a federal legal order built freedom of establishment[318] and subsidiarity.[319] A string of European Court of Justice decisions

---

306 L Terence Blackburn, 'The Societas Europea: The Evolving European Corporation Statute' (1993) 61 Fordham Law Review 695, 697.
307 Paul L Davies, 'Workers on the Board of the European Company' (2003) 32(2) Industrial Law Journal. 75; Karol Linmondin, 'The European Company (*Societas Europaea*) – A Successful Harmonisation of Corporate Governance in the European Union?' (2003) (15)(1) Bond Law Review 147.
308 Paul L Davies (n307) 77;Theo Raaijmakers, 'The Statute for a European Company: Its Impact Upon Board Structures, and Corporate Governance in the European Union' (2004) 5(1) European Business Organisation Law Review 159, 165.
309 SE Employee Involvement Directive (n305) Art 7; Paul L Davies (n307) 90.
310 Paul L Davies (n307) 87.
311 Marios Bouloukos, 'The Legal Status of the European Company (SE): Towards a European Company "a la carte"?' 2004(4) International Business Law Journal 489, 490.
312 SE Employee Involvement Directive (n305) Art 7(2)(a).
313 SE Employee Involvement Directive (n305) Art 7(2)(b); Paul L Davies (n307) 92-4.
314 ibid.
315 Tim Veen, 'The European Company: Overview and Outlook' 2007 50 European Newsletter 5.
316 Wolf Sauter (n302).
317 Stefan Grundmann, 'The Structure of European Company Law: From Crisis to Boom' (2004) 5(4) European Business Organisation Law Review 601, 605.
318 TFEU (n304) Arts 49, 54.
319 Consolidated Version of the Treaty of the European Union [2008] OJ C115/13 ('TEU'), Art 5.

from 1999 onwards has defined the extent of the right to freedom of establishment.[320] A national of State A may register a company in State B and trade mainly, or entirely, in State A while taking advantage of the rules of company law of State B.[321] State A must recognise the legal capacity of such companies[322] regardless of the provisions of its national law.[323] Further, State A cannot discriminate against companies formed in this manner through national law.[324] This competitive process is acknowledged, and to an extent constrained by the Cross Border Merger Directive, which imposes protections for employee participation in the same vein as the SE. [325] However, the constraints of the directive only affect mergers between existing companies and cannot affect companies which liquidate and re-establish elsewhere. Nor is it clear how a German style system of codetermination will fair when transposed into an unsympathetic jurisdiction such as Britain. Therefore, jurisdictional competition is a legal reality. Indeed, it is borne out by the wave of German companies that have moved to Britain in recent years, precisely to avoid codetermination and minimum capital requirements.[326]

While regulatory competition is inevitable, to what extent is it a satisfactory means of attaining convergence? Opinions are divided. The 'race to the bottom' argument suggests that one state, such as Delaware in the US, achieves dominance in the incorporations market by legislating lax rules that permit managers to maximise personal returns at the expense of shareholders.[327] However, this argument both ignores the discipline of the employment, capital, credit and takeover markets and the historical examples of Arizona and South Dakota, both of which over liberalised their corporate law and became disreputable jurisdictions.[328] Based on these counterarguments 'race to the top' theorists argue that corporate constituents are rational actors and would avoid over-liberalised jurisdictions and the effect of market disciplines would result in corporate migration to new jurisdictions which would result in an increase in value.[329] Therefore, they argue that Delaware's success was more attributable to a corporate law that maximises, rather than minimises shareholder value.[330] Other scholars suggest

[320] *Centros Ltd v Erhvervs-og Selskabsstyrelsen* (Case C-212/97) [1999] ECR I-1459.
[321] ibid paras 17, 26-7.
[322] TFEU (n304) Arts 49, 54.
[323] *Überseering BV v Nordic Construction Co Baumanagement GmbH (NCC)* (Case C-208/00) [2002] ECR I-9919.
[324] *Kamer van Koophandel en Fabrieken voor Amsterdam v Inspire Art Ltd* (Case C-167/01) [2003] I-10155 para 105.
[325] Council Directive 005/56/EC of 26 October 2005 on cross-border mergers of limited liability companies ('Cross Border Merger Directive') [2005] OJ L 310/1, Arts 10, 16.
[326] Simon Deakin, 'Is Regulatory Competition the Future for European Integration?' (2006) 13 Swedish Economic Policy Review 13, 73, 83; Gerrit Wiesmann, 'German Companies Flee to the UK' *Financial Times* (London, 24 May 2006); Volker Triebel and Christopher Horton, 'Will More English PLCs Take Off in Germany?' International Financial Law Review (London, 1 July 2006).
[327] William Cary, 'Federalism and Corporate Law: Reflections Upon Delaware' (1974) 83 Yale Law Journal 663.
[328] Didier Martin and Forrest G Alogna, 'A European Delaware: The Nascent Regulatory Market in Europe' (2007) Corporate Finance & Capital Markets Law Review (forthcoming) available at < http://ssrn.com/abstract=1345005> accessed 12 August 2012 19-20, 25.
[329] Daniel R Fischel, 'The "Race to the Bottom" Revisited: Reflections on Recent Developments in Delaware's Corporate Law' (1982) 76 Northwestern University Law Review 913.
[330] ibid 919.

regulatory competition is naturally monopolistic on the basis that there are significant efficiencies to be gained when everyone uses the same system.[331] Therefore, logic, and the US example, suggests a competitive process is likely to eventually result in a *de facto* standard.

In contrast, Charney argues the factors that produce the present state of diversity are unlikely to disappear as a result of this process.[332] It is notable that the European example includes considerably wider historical, cultural, linguistic, governmental and ideological barriers to convergence than the United States. However, in the absence of an ideological consensus or serious legislative solution, it is difficult to conceive of an alternative method.

[331] Michael Klausner, 'Corporations, Corporate Law and Networks of Contracts' (1995) 81 Virginia Law Review 757, 842-7; Didier Martin and Forrest G Alogna (n328) 26.
[332] David Charney, 'The Politics of Corporate Governance' in Jeffrey N Gordon and Mark J Roe (eds), *Convergence and Persistence in Corporate Governance* (CUP 2004) 293, 308-9.

THE OPERATION OF RULES RELATING TO FIDUCIARY RELATIONSHIP – CLEAR, CERTAIN, WORKABLE?

AMY KERR[*]

This essay constitutes the exploration of the rules behind fiduciary duties and accessory liability with respect to their decrease in clarity in English Law. The regulatory function of the rules surrounding fiduciary duties is assessed; particularly with respect to the "no-conflict" and "no profit" principles. In addition, there is also analysis of when there is a breach of duty facilitated by a third party; with a focus on dishonest assistance. It is found that change is needed with regards to these equitable areas of law and a Supreme Court ruling would be welcome to ensure upmost clarity and workability.

While the English courts have been reluctant to define what constitutes a fiduciary relationship,[1] it is clear that its defining characteristic is loyalty.[2] Indeed, when one acts in a fiduciary capacity, one has the potential to affect the legal interests of their principal.[3] For this reason, it is essential that the principal can wholeheartedly rely on the loyalty of the fiduciary.[4] Where the duty of loyalty is breached, the common law has rigidly operated both to compensate the principal and to regulate, and deter, the misconduct of those acting in a fiduciary capacity. However, as the common law rules have progressed, their clarity has arguably deteriorated, particularly where breach of fiduciary duty is facilitated by a third party.

Although fiduciary relationships often arise in the context of commercial agreements, the duties a fiduciary owes his principal are more extensive than contractual obligations.[5] Indeed, because a principal relies on his fiduciary to act in his best interests, he puts himself in a position of vulnerability. Therefore, the common law's insistence on ensuring the fiduciary is unable to misuse his position seems only logical. To this effect, the regulatory function has been a key feature of its development, centring on two fundamental requirements. Firstly, a fiduciary must not put himself in a position where his duty of loyalty and his personal interests may conflict (the 'no conflict' principle), and he must not make any unauthorised profits as a result of his position (the 'no profit' principle).[6] The importance of these principles was reiterated by Lord Justice Millet in *Bristol and West BS v Mothew*[7] where the additional duties of

---

[*] Amy Kerr, Newcastle University, Law LLB Stage Two.

[1] Matthew Conaglen, 'The Nature and Function of Fiduciary Loyalty' [2005] LQR 452; See, for example, *Lloyds Bank Ltd v Bundy* [1975] QB 326, 341.

[2] *Bristol and West BS v Mothew* [1998] Ch 1 18; *Sinclair Investments (UK) Ltd v Versailles Trade Finance Ltd* [2010] EWHC 1614 (Ch).

[3] *Frame v Smith* [1987] 42 DLR (4th) 81.

[4] *Bristol and West BS v Mothew* [1998] Ch 1, 18.

[5] *Re Goldcorp Exchange Ltd* [1995] 1 AC 74 *per* Lord Mustill; Matthew Conaglen, 'The Nature and Function of Fiduciary Loyalty' [2005] LQR 452.

[6] *Bray v Ford* [1896] AC 44.

[7] [1998] Ch 1, 18.

acting in good faith,[8] and not to the benefit or himself or a third party (except with the express consent of the principal) were added to the list of fiduciary duties.

Let us consider the 'no conflict' principle, which has been 'established as the irreducible core of fiduciary obligation'.[9] Indeed, the principle is strictly applied to prevent fiduciaries from furthering their own interests at the expense of the vulnerable principal[10] to the extent that even the possibility of conflict is sufficient to constitute a breach of duty.[11] Certainly, the principle aims to make it impossible for fiduciaries to take personal benefit from their position[12] by ensuring that where this happens, the principal is entitled to rescission or an account of profits, stripping the fiduciary of any gain.[13] This is a clear and settled principle, illustrated, for example, by the rule laid down in *Keech v Sandford.*[14] This rule prevents a fiduciary from renewing a lease for his own benefit, which he was able to obtain solely from his position as trustee to the original lease, even where he has unsuccessfully attempted renew the original lease on behalf of the beneficiary.[15] This case clearly demonstrates that the common law strictly applies the 'no conflict' principle even where the result seems harsh. However, on consideration of the nature of the remedies available following breach of the no conflict principle, it is clear that they are regulatory rather than punitive,[16] and as such, perhaps its inflexibility is justifiable.

Another area in which the case law demonstrates its reluctance to relax the rules for breach of fiduciary duty is where the fiduciary makes an unauthorised profit from his position. Indeed, it is clear that where the fiduciary makes a secret profit or takes a bribe, the money will be held on constructive trust for the principal.[17] However, there is ambiguity regarding whether the appropriate remedy should be proprietary or personal. In *Lister v Stubbs*[18] for example, a personal remedy was deemed appropriate after the fiduciary duty was breached by taking bribes. This entitled the principal to recover the value of the bribe but not any profit accruing from it. However, in the more

[8] However, cf Matthew Conaglen, 'The Nature and Function of Fiduciary Loyalty' [2005] LQR 452, 456.
[9] Ernest J Weinrib, 'The Fiduciary Obligation' [1975] 25 UTLJ 1, 16.
[10] See *Bray v Ford* [1896] AC 44, 51; *Aberdeen Railway Co v Blaikie Bros* (1854) 1 Macq 461, 471;Matthew Conaglen, 'The Nature and Function of Fiduciary Loyalty' [2005] LQR 452*;* Tsun Hang Tey, 'Fiduciaries, third parties and remedies - Singapore's perspectives and contribution' [2010] TLI 234.

[11] *Regal (Hastings) Ltd v Gulliver* [1967] 2 AC 134; *Boardman v Phipps* [1967] 2 AC 46; Tsun Hang Tey, 'Fiduciaries, third parties and remedies - Singapore's perspectives and contribution' [2010] TLI 234.
[12] See *Boardman v Phipps* [1967] 2 AC 46; Alastair Hudson, 'Recent cases suggesting moving away from Boardman v Phipps', available at <http://www.alastairhudson.com/trustslaw/Recent%20cases%20suggesting%20moving%20away%20from%20Boardman%20v%20Phipps.pdf> accessed 21 January 2012.
[13] ibid, see also *Docker v Somes* (1834) 39 ER 1095.
[14] [1726] Sel Cas Ch 61.
[15] ibid; Hanbury & Martin, *Modern Equity* (18th edn, Sweet & Maxwell 2009) 639.

[16] Matthew Conaglen, 'The Nature and Function of Fiduciary Loyalty' [2005] LQR 452, 465.
[17] *AG of Hong Kong v Reid* [1994] 1 AC 324.
[18] [1890] 45 Ch D 1.

recent case of *AG of Hong Kong v Reid*[19] the Privy Council awarded a proprietary remedy, after the bribe money in question was used to purchase a house which had increased in value. Indeed, it was held that both the original bribe and the profit from it were to be held on constructive trust, as Lord Templeman suggested that it would be unconscionable for the fiduciary to retain any benefit resulting from his breach of duty and, as such, the profit should compensate the wronged party.[20] Yet the approach taken by the Court of Appeal in *Lister v Stubbs* has not been overruled and, as such, remains binding. The decision of the Privy Council is merely persuasive and that the Court of Appeal in *Sinclair Investments (UK) Ltd v Versailles Trade Finance Ltd*[21] followed the approach taken in *Lister* suggests that future cases will be decided similarly. However, the law in this area is not entirely certain. The *obiter* comments of Lord Neuberger MR in *Sinclair Investments* suggest that personal remedies are awarded because the doctrine of precedent makes the decisions in *Lister,* and the subsequent cases binding, not because the decision in *AG of Hong Kong v Reid* was incorrect.[22] This is an example of English law being burdened by its 'historical baggage'[23] and, indeed, should the decision in *Lister* be overruled, it is possible that proprietary remedies will be awarded in the future.[24]

An additional aspect of the 'no profit' rule which lacks clarity is where the fiduciary acts in good faith and benefits the principal, yet at the same time makes a profit. Such was the case in *Boardman v Phipps,*[25]where the majority of the House of Lords upheld the claim that the personal profit made by Mr Boardman as trustee was property of the trust. This was because the information which facilitated the profit was obtained in his fiduciary capacity. He had therefore purported to represent the trust, making the information received, and the subsequent profits, trust property.[26] Notably, it was sufficient that Boardman had put himself in a position where his duty and personal interests could conflict to require him to account for his profits, despite that fact no conflict occurred.[27] Not only was this decision not unanimous, the controversy of its application became apparent in *Murad v Al-Saraj*[28]where the court expressed doubt about its harshness and inflexibility. Certainly, Lady Justice Arden highlighted that perhaps it was now necessary for the law to be relaxed in situations where the fiduciary

[19] [1994] 1 AC 324.
[20]*Sinclair Investments (UK) Ltd v Versailles Trade Finance Ltd* **[**2010] EWHC 1614, 51; *AG of Hong Kong v Reid* [1994] 1 AC 324; '*The Ghost of Lister v Stubbs'* available at <http://lawbore.net/articles/Lister_Stubbs.htm> accessed 23 January 2012.
[21] **[**2010] EWHC 1614.
[22] ibid 73 *per* Lord Neuberger MR.
[23] See Tsun Hang Tey, 'Fiduciaries, third parties and remedies - Singapore's perspectives and contribution' [2010] TLI 234.
[24] ibid; *Ultraframe (UK) Ltd v Fielding* [2005] EWHC 1638 (Ch) para 1344.
[25] [1967] 2 AC 46.
[26] *Boardman v Phipps* [1967] 2 AC 46 *per* Lords Hodson and Guest; Oakley, *Constructive Trusts* (3rd edn, Sweet & Maxwell 1997) 168-169.

[27] *Boardman v Phipps* [1967] 2 AC 46 *per* Viscount Dilhorne; Tsun Hang Tey, 'Fiduciaries, third parties and remedies - Singapore's perspectives and contribution' [2010] TLI 234.
[28] [2005] EWCA Civ 959.

acts in good faith and to the benefit of the beneficiary.[29] It is therefore unclear whether the decision in *Boardman* will retain its validity in future cases. A potential justification for the stern approach in *Boardman* is that it is in line with the strict approach concerning the 'no conflict' principle. After all, the 'no profit' principle has been regarded as merely an extension of the 'no conflict' principle;[30] and if this is taken to be correct perhaps the inflexibility of the common law approach can be justified in terms of consistency.

The law's clarity and workability deteriorates further when one examines the case law in which breach of duty has been facilitated by someone outside the fiduciary relationship. Third party liability for breach of fiduciary duty can arise where a third party assumes the role of fiduciary,[31] dishonestly assists the fiduciary, or knowingly receives property resulting from breach of duty. The rules regarding what constitutes both dishonest assistance and knowing receipt are far from clear, certain and workable;[32] however, this article considers the lack of clarity surrounding dishonest assistance only.

Lord Nicholl's speech in the Privy Council case of *Royal Brunei Airlines v Tan*[33] was initially welcomed for clarifying and simplifying the requirements for imposing secondary liability for dishonest assistance.[34] Certainly, removing the 'anomalous' requirement[35] that the fiduciary's breach be fraudulent appeared to address the loophole resulting from the rules laid down in *Barnes v Addy*,[36] while replacing the problematic requirement of the 'knowledge'[37] with an objective test of dishonesty seemed to address the issue of how the accessory's state of mind was to be assessed.[38] Indeed, Lord Nicholls clearly stated that dishonesty simply meant 'not acting as an honest person would in the circumstances'[39] and that it was to be assessed using an objective standard.[40] Nonetheless, the House of Lords in *Twinsectra Ltd v Yardley*[41] appeared to

[29] ibid para 82; see also Alastair Hudson, 'Recent cases suggesting moving away from Boardman v Phipps', available at <http://www.alastairhudson.com/trustslaw/Recent%20cases%20suggesting%20moving%20away%20from%20Boardman%20v%20Phipps.pdf> accessed 21 January 2012.
[30] *Swain v The Law Society* [1983] 1 AC 598; *Boardman v Phipps* [1967] 2 AC 46.
[31] *Mara v Browne* [1896] 1 Ch 199.

[32] Charles Mitchell, 'Dishonest assistance, knowing receipt, and the law of limitation' [2008] Conv 226.
[33] [1995] 2 AC 378.
[34] Nikunj Kiri, 'Recipient and accessory liability- where do we stand now?' [2006] JIBLR 611.
[35] ibid; *Royal Brunei Airlines v Tan* [1995] 2 AC 378, 385; Hans Tjio, 'No Stranger to Unconscionability' (2001) JBL 299.
[36] (1873-74) LR 9 Ch App 244; *Baden, Delvaux and Lecuit v Société Générale pour Favoriser le Développement du Commerce et de l'Industrie en France SA* [1992] 4 All ER 161.
[37]It was unclear whether actual, constructive or imputed knowledge was required – see *Eagle Trust plc v SBC Securities Ltd* [1992] 4 All ER 488; cf *Polly Peck International v Nadir (No 2)* [1992] 4 All ER 769.
[38] *Royal Brunei Airlines v Tan* [1995] 2 AC 378, 389; Hans Tjio, 'No Stranger to Unconscionability' [2001] JBL 299.
[39] *Royal Brunei Airlines v Tan* [1995] 2 AC 378, 389.
[40] ibid.
[41] [2002] 2 AC 164.

alter the requirements by introducing a combined test of dishonesty, comprising of both objective and subjective elements. As under *Twinsectra*, liability was dependent both upon the defendant's conduct being classed as dishonest by honest people, and, crucially, upon the defendant's acknowledgement that honest people would class his behaviour as dishonest. This decision was highly contentious. Indeed, not only did it depart from Lord Nicholl's judgement, but it set too high a threshold for imposing liability for dishonest assistance.[42] The result was a 'retreat' from *Twinsectra* towards the preferred *Royal Brunei*; however, without the decision of the House of Lords being overruled, the outcome was that of judicial decisions which lacked in clarity and consistency.[43]

Although the Privy Council in *Barlow Clowes International Ltd (In Liquidation) v Eurotrust International Ltd*[44] affirmed that the test was to be objective, their justification that the decision in *Twinsectra* was 'misinterpreted' is hardly convincing.[45] Indeed, one must consider that the dissenting comments of Lord Millet, who rejected the subjective element as unnecessary, were ignored. It is therefore difficult to conclude that Lords Steyn and Hoffman, who were in the majority in *Twinsecta,* were now purporting to agree with Lord Millet's dissent in *Barlow Clowes*. In any case, the clarification provided marked a significant judicial change in policy[46] which was echoed by the Court of Appeal in *Abou-Rahmah v Abacha.*[47] The case is particularly significant, as for the first time it was expressly stated that the majority decision in *Twinsectra* had been misinterpreted, and did not in fact introduce the aforementioned subjective requirement.[48] Further, the Court of Appeal's endorsement of the decision in *Barlow Clowes* removed ambiguity over which approach should be taken in future cases, as the decision in *Barlow Clowes* as a Privy Council decision was merely persuasive authority.[49] However, it is interesting to note the judicial make-up of the Pricy Council in *Barlow Clowes,* as at the time all their Lordships were members of the appellate committee of the House of Lords.[50] Therefore, it is 'difficult to see that another constitution of the Appellate Committee would itself come to a different view'.[51]

By removing the subjective requirement, the circumstances in which dishonest third parties can escape liability for participation are limited, making the rules for dishonest

---

[42] See Alistair Hudson, 'Liability for dishonest assistance in a breach of fiduciary duty,' available at <http://www.alastairhudson.com/trustslaw/DAMar07.pdf> accessed 21 January 2012.
[43] See Desmond Ryan, 'Royal Brunei dishonesty: a clear welcome from Barlow Clowes' [2007] Conv 168.
[44] [2006] 1 WLR 1476.
[45] Desmond Ryan, 'Royal Brunei dishonesty: clarity at last?' [2006] Conv 188, 191, 194-195.
[46] Nikunj Kiri, 'Recipient and accessory liability- where do we stand now?' [2006] JIBLR 611.
[47] [2006] EWCA Civ 1492.
[48] ibid 65 *per* Arden LJ.
[49] Desmond Ryan, 'Royal Brunei dishonesty: a clear welcome from Barlow Clowes' [2007] Conv 168, 172.
[50] ibid 173.
[51] ***Abou-Rahmah v Abacha*** [2006] EWCA Civ 1492, 68 *per* Arden LJ; Desmond Ryan, 'Royal Brunei dishonesty: a clear welcome from Barlow Clowes' [2007] Conv 168, 173.

assistance in line with the strict approach favoured by the common law for breach of duty by the fiduciary himself. Justifiably, such a subjective element is necessary in the criminal law; however, dishonest assistance as a secondary form of civil liability need not take account of the defendant's views on the morality of his actions.[52] What constitutes dishonest assistance is now reasonably clear; however, a decision of the Supreme Court specifying the precise requirements would remove any lingering uncertainty. There was scope for this opportunity to arise in *OBG v Allan*[53] and it is disappointing that it was not utilised to provide clarification once and for all.[54]

What is certain is that the English law refuses to tolerate any breach of fiduciary duty and its harsh approach serves to deter disloyal fiduciaries, and those who assist in or benefit from such a breach. However, it is possible that the rigid application of the 'no profit' principle will be relaxed in situations where the fiduciary acts in good faith and to the principal's benefit yet receives a personal profit. Arguably such a relaxation is to be welcomed for being a more equitable approach which would prevent unfortunate decisions like that of *Boardman v Phipps.* In terms of the future for dishonest assistance, a ruling of the Supreme Court would be invaluable in enhancing the clarity and workability of the law regarding the test for dishonesty. However, the clarity provided by the Court of Appeal in *Abou-Rahmah v Abacha* is highly significant, and the likelihood is that a ruling from the Supreme Court would simply echo this judgement.[55]

[52] Desmond Ryan, 'Royal Brunei dishonesty: a clear welcome from Barlow Clowes' [2007] Conv 168, 173.
[53] [2008] 1 AC 1, 75, 90.
[54] ibid; Tsun Hang Tey, 'Fiduciaries, third parties and remedies - Singapore's perspectives and contribution' [2010] TLI 234, 239.

[55] Desmond Ryan, 'Royal Brunei dishonesty: a clear welcome from Barlow Clowes' [2007] Conv 168, 173.

PIRATES AHOY! COPYRIGHT AND INTERNET FILE-SHARING

*ALAN ROYLE**

This paper investigates how the digital revolution has changed the way in which we access copyright protected material and whether the law has evolved to protect this. The paper takes a balanced approach to the issue of online file-sharing, evaluating the effects it has had on the entertainment industry. Furthermore it discusses whether online sharing can co-exist with the current copyright system. It concludes that file-sharing networks themselves are useful tools which promote creativity and that strict legislation would in fact be harmful to the industry. The Digital Economy Act 2010 is examined in detail and it is the opinion of the author that it will do little to reduce piracy and will place unnecessary burdens on internet service providers. It concludes that file-sharing networks can co-exist with the current copyright regime. In addition it is suggested that further legislation such as the Digital Economy Act would be counter-productive. Alternatively it is presented that it is the responsibility of the entrenched industries to combat privacy. This should be done by evolving with technology and offering viable alternatives to consumers.

1 INTRODUCTION

In the digital age, piracy is rife on the high seas of the internet. Copyright infringement takes place daily on a massive scale.[1] Vast libraries of music, films and television are available within minutes and completely free of charge.[2] This is truly one of the most important cultural revolutions of recent times. Yet, it is also argued that file-sharing has become so prominent that entertainment industries are facing a catastrophe which will destroy them.

Copyright law has changed very little since its inception with the Statute of Anne 1710.[3] However, the digital revolution has changed this. Technology has stretched the ambit of copyright law to breaking point, with many arguing that the current model of copyright law is not fit for purpose. Others argue that it is the enforcement of copyright law which needs to be strengthened; by giving rights holders more tools to prevent infringement in the digital arena. This thesis will attempt to elucidate the many tensions and conflicting interests present in the current copyright regime. The ultimate goal of the thesis will be to determine whether file-sharing can co-exist with the current copyright regime, without the need for further enforcement legislation.

* Alan Royle, Newcastle University, Law LLB Stage Three.

[1] Ian Hargreaves, 'Digital Opportunity – A Review of Intellectual Property and Growth' (2011) 69 <http://www.ipo.gov.uk/ipreview-finalreport.pdf> accessed 30 April 2012.

[2] Duke, 'Are You a Pirate? – The Vast Scope of Copyright' (Legal Piracy, 1 April 2011) <http://legalpiracy.wordpress.com/2011/04/01/are-you-a-pirate/> accessed 30 April 2012.

[3] Nick Rose and Michael Sweeney, 'The Hargreaves Report' (2011) 22(7) Ent LR 201, 201.

Section 2 of the thesis will discuss how file-sharing infringes copyright. It will examine the purpose of copyright law and the effect of file-sharing on this purpose. Also, the section will deliberate whether file-sharing is truly responsible for the decline of the entrenched entertainment industries, or if there are other factors in play. Section 3 and 4 will discuss the most recent development in UK digital copyright law – the Digital Economy Act 2010 (the DEA). This piece of legislation was a response to lobbying from the entertainment industries and seeks to give rights holder's powerful tools to enforce copyright. Section 5 will examine alternative solutions to tackling piracy, namely, industry cooperation or a non-commercial use levy.

Overall, this thesis seeks to take a balanced approach towards file-sharing. In this author's view it is not acceptable for all copyright in the digital arena to be waived (a state of 'digital abandon')[4] or for all copyright to be locked and protected by, for example, digital rights management (DRM) which would result in a state of 'digital lockup'.[5] Creators and rights holders must receive remuneration for their efforts without stifling internet freedom. There are many different ways to achieve this. This thesis will be seek the most sensible, proportionate and balanced solutions.

2 PROBLEM WITH PIRACY

2.1 Introduction

Copyright law in the digital age is under siege. In the past, infringing the right to copy was a physical process which involved investment in time and materials, and resulted in a physical copy. In the digital age, with all manner of works in digital format, from films to eBooks, copying is almost instantaneous with no degradation of quality. These copies can be distributed across the globe at little or no cost via file-sharing networks. This section will briefly examine how file-sharing infringes copyright. It will also examine the relationship between file-sharing and the purpose of copyright law to attempt to determine whether file-sharing is as damaging as its opponents claim.

2.2 The purpose of copyright law

Copyright law is traditionally said to have two main purposes. First, from a natural law approach, it is designed to secure just and fair reward for works which are part of the author's personality. This natural or moral justification for copyright law is more prominent in continental copyright systems where the author is held in high esteem.[6] Second, and perhaps more pertinent in common law legal systems, is the instrumentalist approach or the 'incentive theory'. It is argued that the protection that copyright law affords should be no more 'than that required to provide a sufficient possibility of return

[4] Mark Nadel, 'Why Copyright Law May Have a Net Negative Effect on New Creations: The Overlooked Impact of Marketing' (Social Science Research Network Electronic Library, 8 January 2003) <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=322120&> accessed 30 April 2012.
[5] Michael Pendleton, 'The Digital Divide - International Enforcement of Digital Lockup' (2006) 1 JILT.
[6] Lionel Bently and Brad Sherman, Intellectual Property Law (3rd edn, OUP 2009) 242.

to induce the creation and dissemination of new works.'[7] This justification is based on the simple assumption that the greater protection afforded to works, the easier it is for authors to secure remuneration for their works, subsequently encouraging them to create more. Therefore, copyright protection appears *prima facie* to serve an important purpose, not just for authors, but for society as a whole by being the engine for the creation of new works. Moreover, it would appear that illegal file-sharing, as it undermines copyright protection, is inherently damaging to society and to our cultural heritage. The relationship between file-sharing and the purpose of copyright will be discussed later in the section.

2.3 Internet file-sharing

Since the late 1990s file-sharing has become increasingly popular. The exponential increase in internet speed, from dial-up to broadband and now fibre-optic, has facilitated the downloading of every type of media. This ranges from individual songs to massive Blu-ray quality films and television shows. The most popular method of file-sharing is by peer-to-peer networking (P2P).[8] P2P networks consist of computers which are linked together over the internet. The P2P software allows computers to communicate and their users (or peers) to search for, access, download and upload material stored in shared folders on the peer's hard drive.[9] The act of downloading a copyrighted work via P2P infringes several rights and could result in liability for the downloader, uploader, P2P network operator and the internet service provider (ISP). As ISP liability is a large and controversial subject, it will only be examined briefly in Section 3 and 4.

The uploader, by putting copyright protected files in the shared folder which can be accessed by other peers on the P2P network, infringes the copyright holder's exclusive right to issue copies to the public. This contravenes s.16 and s.18 Copyright Designs and Patent Act (CDPA). This also infringes the s.20(2) CDPA right of communication to the public by electronic transmission. The downloader, by transferring a file across the internet from a peer, creates a new copy on his or her hard drive. This infringes the copyright holder's exclusive right to make copies (s.16-17 CDPA). The P2P network operator can be held liable under s.16(2) CDPA which provides that 'copyright in a work is infringed by a person who without the licence of the copyright owner does, or authorises another to do, any of the acts restricted by the copyright.' Indeed, a High Court decision recently held that The Pirate Bay, a website which provides .torrent files containing the information needed to download from peers, authorises copyright infringement.[10]

---

[7] Neil Netanel, 'Copyright in a Democratic Society' (1996) 106 Yale LJ 283, 367.
[8] Hendrik Schulze and Klaus Mochalski, 'Internet Study 2008/2009' (ipoque, 2009) <http://www.ipoque.com/sites/default/files/mediafiles/documents/internet-study-2008-2009.pdf> accessed 30 April 2012.
[9] Haflidi Kristjan Larusson, 'Uncertainty in the scope of copyright: the case of illegal file-sharing in the UK' (2009) 31(3) EIPR 124, 124.
[10] Editorial, 'High Court rules The Pirate Bay operators and users guilty of copyright infringement' (Out-Law.com, 8 February 2012) <http://www.out-law.com/en/articles/2012/february/high-court-rules-the-pirate-bay-operators-and-users-guilty-of-copyright-infringement/> accessed 30 April 2012.

It must also be noted that while P2P file-sharing is by far the most popular form of file-sharing (accounting for some 40-70% of internet traffic by volume in 2009), it is not the only method.[11] Usenet is similar to P2P in function but differs in infrastructure.[12] File hosting is also common, through sites such as Rapidshare and MegaUpload. Copyrighted files are uploaded to these websites, and then the download links are distributed via forums or search engines to those who wish to download the file directly from the hosting company.[13] It should be noted that these hosting companies sometimes charge monthly fees to allow users to download at the fastest rate. Some consumers are willing to pay to access pirated content. While this section will focus mostly on P2P file-sharing, it must be remembered that there are many other methods of file-sharing and many more will be developed in the future. Consequently, it is dangerous for the law to remain static and attempt to tackle only P2P file-sharing. Technology moves faster than the law and therefore, a future-proof approach is needed to deal with piracy and not short-term measures which can easily be circumvented.

2.4 The effects of internet file-sharing

There has been much debate over the past decade as to the effects of internet file-sharing. This section will consider, firstly, the effects that internet file-sharing are said to have had on the entertainment industry, primarily the music industry. Secondly, the effects on copyright law as a whole will be considered – does illegal file-sharing undermine the purpose of copyright law? The answers to these questions will determine whether the current copyright regime is fit for purpose in the digital age.

Prima facie, it would seem logical that illegal file-sharing harms the entertainment industry by providing consumers with products, for free, which they would otherwise have to pay for. However, the evidence is conflicting – or at least inconclusive. Oberholzer and Strumpf concluded in 2004 that file-sharing had 'an effect on sales which is statistically indistinguishable from zero'.[14] A Japanese study from the same year found 'very little evidence' that file-sharing reduced record sales.[15] Indeed, despite a continual increase in the number of illegal file-sharers from 1999 to 2004, 2004 was a record year for album sales, with figures up 3% from the previous year.[16] However, there are also studies which argue that file-sharing is having a near catastrophic effect on music sales. The BPI's Digital Music Nation report in 2010 claims that illegal file-

[11] Schulze (n 3).
[12] Ernesto, 'MPAA Lists "Notorious" Pirate Sites To U.S. Government' (TorrentFreak, 28 October 2011) <http://torrentfreak.com/mpaa-lists-notorious-pirate-sites-to-u-s-government-111028/> accessed 30 April 2012.
[13] ibid.
[14] Felix Oberholzer and Koleman Strumpf, 'The Effect of File Sharing on Record Sales An Empirical Analysis' (UNC.edu, March 2004) <http://www.unc.edu/~cigar/papers/FileSharing_March2004.pdf> accessed 30 April 2012.
[15] Tastuo Tanaka, 'Does file-sharing reduce CD sales? A case for Japan' (Hitosubashi University Institute of Innovation Research, December 2004) <http://www.iir.hit-u.ac.jp/iir-w3/file/WP05-08tanaka.pdf> accessed 30 April 2012.
[16] Editorial, 'UK music sees record album sales' (BBC News, 26 November 2004) <http://news.bbc.co.uk/1/hi/entertainment/4044303.stm> accessed 30 April 2012.

sharing cost the UK music industry £219 million in 2010.[17] The Hargreaves Review cites the same study to demonstrate the conflicting evidence on such an important social and economic issue. The BPI report claims that in 2010 65% of music downloads in the UK were illegal, whereas MidemNet's 2010 Global Music Study provides a figure of 13%.[18] Moreover, the Hargreaves Review failed 'to find a single UK survey that is demonstrably statistically robust'.[19] There is no doubt that illegal file-sharing is taking place on a large scale. There is doubt however, as to the damage it is causing entertainment industries.

Besides evidential difficulties, there are also logical difficulties. The proposition that every illegal download causes economic damage to the rights holder, because the downloader would have otherwise purchased the copy (a 'lost sale') is a false dichotomy. It is probable that many copyrighted works, if unavailable for illegal download, would not result in a sale. This difficulty in predicting consumer behaviour also contributes to evidential problems when determining the effects of piracy, as it breaks the causal link between illegal downloads and economic damage to the rights holder.

Moreover, there are other reasons why music sales have seen a general decrease of the past few years, such as economic recession at the beginning and end of the last decade.[20] It could be argued that the strongest explanation is that records, as a sink for disposable income, are facing increased competition from other media. DVDs, eBooks, video games and smart phone apps have become prominent forms of entertainment, especially among young consumers.[21]. It is submitted that while piracy has most likely contributed to some lost sales, there are many other factors involved which cannot be ruled out for the decrease in revenue for entrenched industries.

If there is no strong irrefutable evidence that file-sharing damages sales, is file-sharing adverse to the overall aim of copyright law? The primary aim of copyright law is to encourage the creation of new works – rewarding the author is the traditional way to achieve this. It could be argued that file-sharing, especially on P2P networks, facilitates this. P2P networks allow important social interaction on a global scale. Robert Danay argues that 'p2p file-sharing networks have become indispensable components of numerous pan-global virtual communities wherein cultural artefacts are both shared and discussed in genre-based chat rooms'.[22] This includes the sharing and discussion of both copyrighted works and non-copyrighted works. Joshua Cohen claims this activity is a 'vital element of communication and an essential tool that people use to understand

[17] Alexandra Duboff, 'BPI Digital Music Nation – pirate wars' (2011) 23(3) Ent LR 85, 85.
[18] Ian Hargreaves, 'Digital Opportunity – A Review of Intellectual Property and Growth' (2011) 69.
[19] ibid.
[20] Robert Danay, 'Copyright vs. Free Expression' (2005) 8 Yale Journal of Law and Technology 32, 54.
[21] Daniel Eran, 'Is Piracy Really Killing the Music Industry?' (RoughlyDrafted, 7 March 2007) <http://www.roughlydrafted.com/RD/RDM.Tech.Q1.07/708F20CD-E67D-45C7-AF95-3E1A6AC07C37.html> accessed 30 April 2012.
[22] ibid 48.

themselves, their society and their place in the world'.[23] It is important to understand that works are not created in a vacuum; authors draw inspiration from other works either consciously or subconsciously. The law on originality recognises this fact by holding that the work should originate from the author, but need not be novel or innovative.[24] Therefore, it is submitted that P2P networks can contribute to the creation of new works by providing authors with a massive domain of experience and inspiration.

Incentive theory dictates that copyright infringement will not encourage creativity, as the tighter the copyright protection, the more works will be created due to the possibility of securing rewards. However, the evidence indicates the contrary. A study, undertaken by Raymond Shih Ray Ku, determined that no correlation could be made between expansion of copyright and an increase in the amount of works created. The study even suggested that 'laws limiting or decreasing copyright protection appear more likely and more consistently to be associated with an increase in the number of new works registered.'[25] This is because copyright protection is a 'double-edged sword' – it does not simply incentivise creation, it can also be a financial barrier to creation. The more stringent the copyright protection, the greater the transaction costs for authors wanting to draw on protected works. So while some protection may be desirable for rightful remuneration, the costs for creating new works resulting from stronger protection will offset any increase in incentives. Therefore, protection which prevents use of or access to cultural networks like P2P networks, could stifle creation of new works. It is argued that digital copyright law should seek to strike a balance between ensuring some remuneration for authors, but at the same time respecting the benefit to society that file-sharing networks produce.

Moreover, as the lack of empirical evidence supporting the proposition may suggest, file-sharing may not be that damaging to the economic interests of authors.[26] Works in circulation in these cultural networks benefit from free promotion and a wider audience; with some studies noting a positive correlation between the amount of pirated content downloaded and the amount of legitimate content consumed.[27] For video games, this may lead to the purchase of the full game for locked content, such as online multiplayer. The importance of securing a wide audience is slowly being recognised by the video game industry as more games are adopting the free-to-play model to secure a wide

[23] Joshua Cohen, 'Common Musical Sense: An Intellectual Call to Arms against the Recording Industry,
Radio Deregulation, and Media Consolidation and their Threat to our National Culture and Democracy' (FITEHOUSE) <http://www.fitehouse.com/images/stories/pdf/common-musical-sense.pdf> accessed 30 April 2012.
[24] University of London Press Ltd. v. University Tutorial Press Ltd. [1916] 2 Ch. 601, 608-609.
[25] Raymond Shih Ray Ku, 'Does Copyright Law Promote Creativity? An Empirical Analysis of Copyright's Bounty' (2009) 62 Vand. L Rev 1669, 1673.
[26] Danay (n 15) 55.
[27] Mike Masnick, 'Industry Suppressed Report Showing Users Of Shuttered 'Pirate' Site Probably Helped Movie Industry' (techditrt, 21 July 2011)
<http://www.techdirt.com/articles/20110721/04092915191/industry-suppressed-report-showing-users-shuttered-pirate-site-probably-helped-movie-industry.shtml> accessed 30 April 2012.

player-base, from which income is made from advertising or micro transactions.[28] This is an example of how the technology behind piracy has led to innovation and financial boon in an industry it has been accused of damaging.

For musical works, this could lead to the sale of merchandise and concert tickets, which are more of a direct benefit to artists than record sales, which invariably benefit the record label.[29] Business reality in the recording industry means that a very small percentage of artists receive financial gain from record sales – in 1999 0.013% of all albums released accounted for the majority of sales, making the vast majority of albums unprofitable.[30] Production, promotion and marketing costs are subtracted from the artist's share of profits, meaning that a high threshold of sales must be passed before an artist begins to receive payment. Consequently, it is submitted that if file-sharing did have a negative effect on sales, the vast majority of artists would not be impacted. Moreover, some artists are now embracing the P2P network model of widespread online distribution. In 2007 Radiohead released the album *In Rainbows* online, independent of a record label, and allowed customers to pay whatever they wanted for the album or download it for free. It is estimated that the band could have made up to $10m in initial album sales.[31] This is an example of how the culture of free music distribution has aided a band in circumventing the noxious recording industry, which it previously would have been completely dependent upon. It is submitted that file-sharing does not cause substantial financial harm to the creators of works, and in some cases it may even be a boon.

Therefore, it is submitted that copyright law which seeks to suppress illegal file-sharing can be harmful. As file-sharing networks are a cultural transmission of ideas and expression, communication over these networks are protected by Art.10(1) European Convention on Human Rights (ECHR). As a qualified right, freedom of expression must be balanced against the intellectual property rights of others. Harsh copyright law could not only stifle creativity, but stifle freedom of expression, which is adverse to the public good. Harsh copyright protection and enforcement can also be damaging to the industries who seek to use it protect their financial interests. These industries must recognise that the individuals and communities who download and distribute their works are potential customers. Those who have illegally downloaded the work are clearly part of a market which the industries should seek to engage with; a point which will elaborated upon in Section 5. Prosecuting potential customers can have severe negative effects for the rights holder.

---

[28] Brad Hilderbrand, 'Blizzard may be the latest publisher to embrace Free-to-Play' (GameZebo, 24 February 2012) <http://www.gamezebo.com/news/2012/02/24/blizzard-may-be-latest-publisher-embrace-free-play> accessed 30 April.
[29] Danay (n 15) 56.
[30] ibid 57.
[31] Eliot Van Buskirk, 'Estimates: Radiohead Made Up To $10 Million on Initial Album Sales (Updated)' (Wired, 19 October 2007) <http://www.wired.com/listening_post/2007/10/estimates-radio/> accessed 30 April 2012.

The video games industry provides an example of these negative effects. In recent years, certain publishers have decided to tackle piracy by protecting their games with DRM. For customers who paid for the game, this DRM would be automatically and mandatorily installed on their PC with the game. One type of DRM, used by publisher Ubisoft, requires a constant internet connection to send packets to Ubisoft's server to confirm that the copy of the game was genuine. This means that customers who had paid for the game would not be able to access the game if they have no internet connection; their connection is faulty or even if there is a problem with Ubisoft's verification servers.[32] The irony is that the DRM has not prevented any of the games from being 'cracked' and made available on file-sharing networks without the DRM.[33] Those who decided to pirate the game were not subject to the onerous DRM that paying customers were subjected to. DRM protected products are more expensive to produce and consumers would obviously favour products which do not treat them as presumptive 'criminals'.[34] This is a clear situation where copyright protection goes too far and causes financial and reputational damage to a company.[35] It is therefore submitted that copyright protection must be fair and balanced, not only for the sake of the consumer, but also for the industry.

2.5 Conclusion

File-sharing networks do not undermine the purpose of copyright law. These networks of social interaction and culture provide a commons which can promote creativity. Moreover, the evidence is inconclusive as to the financial harm that file-sharing causes to rights holders, especially to artists. In fact, artists may actually benefit from the infrastructure file-sharing networks provide to disseminate their works. Harsh copyright protection not only limits the benefits of file-sharing, but can also proactively cause financial damage to the entrenched industries by alienating consumers and even encouraging piracy of their products. In practice, harsh enforcement measures may result in even more negative consequences, as discussed in the next section which will examine the DEA as an example of modern anti-piracy law.

3 THE DIGITAL ECONOMY ACT

3.1 Introduction

In August 2009 Lord Mandelson, then Secretary of State for Business, Innovation and Skills, was reported to have dined with Hollywood mogul David Geffen at a villa in

---

[32] Tom Senior, 'Ubisoft DRM server downtime locks players out of Anno 2070, Driver: San Francisco and more' (PC Gamer, 8 February 2012) <http://www.pcgamer.com/2012/02/08/ubisoft-drm-server-downtime-locks-players-out-of-anno-2070-driver-san-francisco-and-more/> accessed 30 April 2012.

[33] Robert Purchese, 'How Bad is PC Piracy Really?' (EuroGamer.net, 30 September 2011) <http://www.eurogamer.net/articles/2011-09-30-how-bad-is-pc-piracy-really-article> accessed 30 April 2012.

[34] Cory Doctrow, 'The DRM Sausage Factory' in The United States Pirate Party (ed), No Safe Harbor: Essays About Pirate Politics (2012) 77.

[35] Rob Zacny, 'Opinion: Ubisoft, piracy, and the death of reason' (PC Gamer, 7 October 2011) <http://www.pcgamer.com/2011/10/07/opinion-ubisoft-piracy-and-the-death-of-reason/> accessed 30 April 2012.

Corfu.[36] The Digital Britain report had been published two months previously, which set out a strategic vision to allow UK to compete in the global digital economy by improving freedom of internet access for the whole country.[37] *The Times* reported that an unnamed Whitehall source had confessed that Mandelson had shown little interest in the Digital Britain agenda before this meeting in Corfu.[38] Yet in the following weeks and months it appeared that Mandelson played a significant role in introducing the most controversial measures into the implementation of the Digital Britain report (the Digital Economy Bill). Mandelson denied the two events were linked.[39] These controversial measures included the possible internet disconnection of copyright infringers and website blocking. The decision to include user disconnection was explicitly opposed to Lord Carter's recommendation in the Digital Britain report.[40]

The first reading of the new Digital Economy Bill was put before the House of Commons on 16 March 2010. After two more readings and very little debate, the bill was granted Royal Assent on 8 April 2010. The bill was passed during the 'wash-up' before the dissolution of parliament when uncontroversial legislation is supposed to be rushed through. The DEA came into force on 12 June 2010.[41]

The DEA was conceived and born in a furore of controversy, which has not relented. This section will analyse the provisions in the DEA and attempt to answer one fundamental question. Will the DEA be effective in dealing with illegal file-sharing?

3.2 Structure of the Digital Economy Act

As stated, the DEA was one of the outcomes of the Digital Britain report which recommended improving the UK's communications infrastructure to, in turn, improve competitiveness in the global digital economy. Most of the 48 sections of the DEA consist of fairly non-contentious provisions pertaining to digital infrastructure. However, ss.3-18 set up the government's new system to deal with online copyright infringement; a system that received no mention in the Digital Britain report. These measures have garnered much criticism from digital rights campaigners and academics, but on the other hand have received support from rights holders and creative industries.

---

[36] Richard Wray, 'Mandelson web cutoff plan 'potentially illegal' (The Guardian, 25 August 2009) <http://www.guardian.co.uk/technology/2009/aug/25/mandelson-web-cutoff-plan-attacked> accessed 30 April 2012.

[37] Department for Business, Innovation & Skills, Digital Britain Final Report (June 2009) <http://www.official-documents.gov.uk/document/cm76/7650/7650.pdf> accessed 30 April 2012.

[38] Jonathan Oliver, 'Mandelson targets web pirates after dinner with mogul' (The Times, 16 August 2009) <http://technology.timesonline.co.uk/tol/news/tech_and_web/the_web/article6797844.ece> accessed 30 April 2012.

[39] Daniel Martin, 'Mandelson launches crackdown on file sharing... just days after meeting with record producer' (Mail Online, 17 August 2009) < http://www.dailymail.co.uk/news/article-1206901/Mandelson-launches-crackdown-file-sharing--just-days-meeting-record-producer.html> accessed 30 April 2012.

[40] Martin (n 4).

[41] Digital Economy Act 2010.

These anti-piracy measures can roughly be divided into three distinct systems. Firstly, ss.3-8 put in a place 'initial obligations' for ISPs to notify infringing subscribers and compile lists of repeat infringing subscribers on behalf of rights holders. These lists can then be used by the rights holders to sue for infringement. Secondly, ss.9-12 introduced 'technical measures' which allow for an escalated approach to dealing with persistent copyright infringers – imposing obligations on the ISPs to limit or even disconnect infringing subscribers' internet connections. Thirdly, ss.17-18 allows for the Secretary of State to block access to an internet location 'from which a substantial amount of material has been, is being or is likely to be obtained in infringement of copyright', or 'made available' or 'used to facilitate access' to material in infringement of copyright.[42] However, the government has recently announced that it scrapped plans to implement website blocking under s.17 and therefore it will not be examined.[43] The implementation and enforcement of the initial obligations and the technical measures fall to Ofcom (ss.6, 11). Ofcom published a draft of the initial obligations code on 28 May 2010.[44] The initial obligations and technical measures will be evaluated below.

3.3 Initial Obligations

The initial obligations set out in ss.3-8 are to function as follows. A rights holder may, for example, search a P2P network for copies of his or her work. Upon finding any copies, the rights holder can take a note of the internet protocol (IP) address which is hosting the copy. The rights holder will be able to discover which ISP the subscriber hosting the copy is using. The rights holder can then send a 'Copyright Infringement Report' (CIR) to the ISP detailing the alleged infringement. The ISP is then obligated to notify the subscriber that they have received information of an alleged infringement. If a subscriber reaches a threshold of CIRs, the ISP is then obligated to provide the rights holder with a 'Copyright Infringement List' (CIL) which details the alleged infringements under the IP address(es) of the anonymous subscriber. This CIL could be used by the rights holder to obtain a court order to reveal the subscriber's identity and bring an action for copyright infringement.

There is one blindingly obvious flaw in this procedure. The assumption – that the IP address gathered by the rights holder indicates that the subscriber is responsible for infringing copyright – is erroneous. There are several reasons for this. Those who are responsible for *en masse* distribution of copyrighted material are unlikely to let their real IP addresses be so readily discovered. Proxies and virtual private networks (VPNs) which mask the user's IP address are easy to obtain and use. Those more determined to hide their identity can even hijack the IP address of an unsuspecting subscriber and use that address for distributing copyrighted material. Subsequently, it has been argued that

[42] s.17(4) Digital Economy Act 2010.
[43] Mark Sweney, 'Government scraps plan to block illegal filesharing websites' (The Guardian, 3 August 2011) <http://www.guardian.co.uk/technology/2011/aug/03/government-scraps-filesharing-sites-block> accessed 30 April 2012.
[44] Ofcom, 'Online Infringement of Copyright and the Digital Economy Act 2010 - Draft Initial Obligations Code' (28 May 2010).

the DEA is targeting the wrong type of internet subscriber.[45] The mainstream internet users who play little or no role in making available copyrighted material will be affected by the DEA's notification system, but those who engage in illegal file-sharing on 'an industrial scale' whilst masking their true IP addresses will not be perturbed by the legislation. Moreover, if a high profile file-sharer did receive a notification letter, it would most likely be ignored or even used as motivation to find a way to effectively mask their IP address. [46]

The targeting of subscribers also illuminates a problem with the inflexibility of the DEA. If the legislation fails to take account of proxies and VPNs, which are relatively old and widely available technologies, how can it expect to function as planned in the future? Legislation dealing with information technology must take account of the innovation in this sector. The provisions in the DEA were barely suitable for use in 2010, never mind in the years to come which will surely bring even more accessible ways to hide one's identity online.

The methods employed by rights holders to identify those who engage in illegal file-sharing of their works have stirred controversy recently. Rights holders will typically employ a law firm who then subcontract to forensic analysts to discover the identities of those who are allegedly infringing their copyright. In 2010 *The Guardian* reported that a couple, aged 56 and 68, received a threatening letter from a law firm demanding compensation for allegedly making a video game available for copying on the internet.[47] It is not difficult to believe their claim that they did not know what P2P file-sharing was. It could be that their connection had been hijacked by an illegal file-sharer or that the method for gathering the infringing IP address simply produced the wrong result. Deborah Finch, head of legal affairs at Which?, stated that 'hundreds and possibly thousands of people have been wrongly identified, on the basis of flawed evidence'.[48]

Presumably, the methods of gathering IP addresses for the CIRs under the DEA could suffer from the same identification problems. In a case brought to court by the notorious firm ACS:Law, who sent out similar threatening letters, Judge Colin Birss QC cast doubt on the legitimacy of using IP addresses to identify particular infringers and heavily criticised the practices of ACS:Law.[49] Under intense scrutiny and an investigation by the Solicitors Regulation Authority, Andrew Crossley, founder of

[45] Nick Cusack, 'Is the Digital Economy Act 2010 the most effective and proportionate way to reduce online piracy?' (2011) 33(9) EIPR 559, 559.
[46] ibid 561.
[47] Dan Sabbagh, 'Digital Economy Act likely to increase households targeted for piracy' (The Guardian, 12 April 2010) <http://www.guardian.co.uk/media/2010/apr/12/digital-economy-bill-households-piracy> accessed 30 April 2012.
[48] ibid.
[49] Media CAT Ltd v Adams & Ors [2011] EWPCC 6.

ACS:Law, announced that the firm would 'close permanently'.[50] Such was the hostile reaction to practices employed. Although the initial obligations under the DEA will remove the practice of law firms sending threatening letters directly to the alleged infringing subscriber, the problem with wrongful identification will remain.

Although a person may not be responsible for distributing infringing material on a particular IP address, the can still be liable under the DEA. Section 3 indicates that notification letters may be sent to subscribers who 'allow infringement'. This means that, for example parents (as subscribers) could be liable for the activity of their children – which is problematic as parents may not have sufficient technical knowledge to prevent this.[51]

It appears that this provision will also affect Wi-Fi providers – but the law is confusing in this respect. Today in the UK, wireless internet is provided in many public places including outdoor spaces. As providers of an internet service, it appears those operating these networks will be under the initial obligations of the DEA – namely, to provide rights holders with CILs. In practice, this is impossible. The Wi-Fi provider cannot 'document and record the personal details of each and every temporary subscriber who takes advantage of the wireless internet access' and to do so would 'negate all the commercial advantages of offering such a service in the first place'.[52]

Moreover, such an obligation is directly contrary to the whole purpose of the Digital Britain recommendations – to improve freedom of internet access across the country. Such obligations would throttle freedom to provide and access wireless internet and would also 'amount to a constructive prohibition on the lawful establishment of Wi-Fi networks'.[53] The DEA seems ignorant of the potential ruinous constraints these obligations could put on businesses which specialise in providing Wi-Fi.

However, because those who provide Wi-Fi normally receive their connection through a larger ISP, it is unclear whether, for the purposes of the act, they would be classed as providers or subscribers. They are technically both. Who should the rights holder send the CIR to? The Wi-Fi provider or the larger ISP? Again, we see the problems with using IP addresses to attempt to identify and assign liability for infringement. The legislation has generated considerable uncertainty regarding obligations for Wi-Fi providers. One can only conclude that these provisions were not thoroughly considered or debated.

---

[50] Kelly Fiveash, 'Patent judge hits out at legal tactics used against file-sharers' (The Register, 9 February 2011) <http://www.theregister.co.uk/2011/02/09/acs_law_media_cat_judgment/> accessed 30 April 2012.
[51] Benjamin Farrand, 'The Digital Economy Act 2010 - a cause for celebration, or a cause for concern?' (2010) 32(10) EIPR 536, 536.
[52] Cusack (n 10) 561.
[53] ibid.

Another flaw in the initial obligations appeared after Ofcom published the Draft Code. They proposed that initially the Code should only apply to ISPs with more than 400,000 subscribers.[54] The argument supporting this decision is that this threshold represents the seven largest ISPs in the UK, which account for approximately 93.4% of the residential and SME broadband market, and so the initial obligations will apply the vast majority of subscribers.[55] Again, the DEA seems to be targeting the wrong type of illegal file-sharer. The committed large-scale file distributer will be under no hardship to switch to a smaller ISP to avoid detection under the initial obligations. It has also been argued that the provision is anti-competitive, because the legislation imposes an incentive for subscribers to cancel their contracts with the larger ISPs if they want their activities online to remain private and unperturbed.[56] Furthermore, s.14 DEA provides a penalty of £250,000 on an ISP which does not comply with the obligations. In short, the DEA puts onerous time-consuming and financial burdens on ISPs, who provide vital infrastructure for 'Digital Britain'. This may result in an increased cost for consumers. It is submitted that ISPs have a strong justification to be angered with the DEA. As a result, they launched a judicial review of the legislation which will be discussed in Section 4.

3.4 Technical Measures

The provisions which have stirred the most heated debate are the 'technical measures' provided in ss.9-12. Once the initial obligations have been in force for at least twelve months and only if they appear unsuccessful in reducing online copyright infringement, the Secretary of State can set in motion plans to impose technical obligations on ISPs. These plans can only come into force after a sixty day consultation period and with the approval of both Houses of Parliament. The DEA does not specify the precise mechanism for the implementation of these measures. Although it may appear that analysis of the technical measures is purely academic for the time being, they are lodged in the legislation like 'Chekhov's gun'. The government felt the need to include these measures as a legitimate means of dealing with piracy.

Although the technical measures will be subject to much consultation, the legislation explicitly defines what they will entail in s.9(3):

> 'A "technical measure" is a measure that—
> a) limits the speed or other capacity of the service provided to a subscriber;
> b) prevents a subscriber from using the service to gain access to particular material, or limits such use;
> c) suspends the service provided to a subscriber; or
> d) limits the service provided to a subscriber in another way.'

[54] Ofcom (n 9) 4.
[55] The Queen (on the Application of), British Telecommunications Plc, Talktalk Telecom Group Plc v The Secretary of State for Business, Innovation and Skills [2011] EWHC 1021, 201.
[56] Cusack (n 10) 562.

Needless to say, these measures have been criticised as draconian. The criticism is even more marked given that problems with identification and vicarious liability have already been established. For example, a whole household could face disconnection because a child downloads copies of MP3s.

In developed countries internet access is a fundamental component of infrastructure. This was recognised in the Digital Britain report itself.[57] It is doubtful that Lord Carter considered throttling or even suspending subscribers' internet access as a way to improve the UK's digital infrastructure. Internet access is important to subscribers for many reasons – for keeping in touch with family and friends, for business, shopping and banking, for educational purposes and cultural interaction as mentioned in Section 2. It is difficult to understand how, in the digital age, the government could even imagine a regime by which homes were disconnected from the internet. It is disturbing that there was little or no protest against these measures by MPs in parliament. Perhaps this can be attributed to the frantic nature of the wash-up period or the recognition that the measures require further consultation.

As demonstrated in Section 2, copyright exists to encourage the creation of new works.[58] In order for new works to be created, authors must have access to the works of others. They must be free to be influenced by, draw from and be inspired by the works of their predecessors because 'no author creates in a vacuum'.[59] Every year sees an increase in the number of musical, literary and dramatic work available on the internet – either by illegal or legitimate means. Therefore, disconnecting homes from the internet could jeopardise cultural progress in the UK.[60] However, it is highly unlikely that the disconnection would ever happen on such a large scale as to threaten cultural creativity, but the argument demonstrates just how far the law has swung in favour of rights holders, at the expense of authorship – the very process that the law was originally designed to encourage.

Internet access is also very important in the information age for another reason. People form and express their opinions and views via the internet. The internet is the modern stalwart of freedom of expression. Perhaps the most extraordinary example of this is how social networking and blogging facilitated the democratic revolutions in the Arab world.[61] It is not surprising that there has been a movement in the United Nations to make internet access a human right.[62] Indeed, recital (4) Directive 2009/140/EC explicitly states interest access is a requirement for the practical exercise of freedom of

[57] Department for Business, Innovation & Skills (n 2).
[58] Jessica Litman, 'The Public Domain' (1990) 39 Emory LJ 965, 969.
[59] Raymond Shih Ray Ku, Does Copyright Law Promote Creativity? An Empirical Analysis of Copyright's Bounty (2009) 62 Vand. L Rev 1669, 1714.
[60] Enrico Bonadio, 'File-sharing, copyright and freedom of speech' (2010) 33(10) EIPR 619, 620.
[61] Kate Taylor, 'Arab Spring really was social media revolution' (TG Daily, 21 September 2011) <http://www.tgdaily.com/software-features/58426-arab-spring-really-was-social-media-revolution> accessed 30 April 2012.
[62] Bonadio (n 27).

expression.[63] In an age when the internet has gained a revered status as the champion of freedom of expression, the UK government has introduced a regime which could see whole families disconnected from the internet. The absurdity of this proposition is even more prominent given that Housing Minister Grant Shapps recently stated that 'internet connectivity should be a necessity, not a luxury'.[64]

Freedom of expression is, of course, protected by Art.10 ECHR. However, it is a qualified right, and can be limited to protect the rights of others. Therefore, the protection of copyright could be a justification for restricting freedom of expression. The restriction, however, must be necessary.[65] It is submitted that internet throttling or disconnection is not necessary to protect copyright. There are alternative ways to enforce copyright which do not result in such an oppressive constraint on the freedom of expression. In any case, because internet disconnection concerns balancing these two important rights, it should be for the courts to make the final decision on disconnection. This was the view taken by the French Constitutional Court reviewing the internet disconnection procedure in the HADOPI legislation.[66] The court held that it was unconstitutional for an administrative agency to have the power to terminate internet access when careful balancing of the two interests at stake was necessary.[67] Therefore, perhaps the most disturbing thing about the technical measures in the DEA is the fact that they do not require a judicial ruling to apply to individual subscribers. There are no references to court orders in ss.9-12 and so it appears that Ofcom will be responsible for having ISPs impose the technical measures on subscribers, and not a court of law, which is a frightening prospect.

3.5 Conclusion

The DEA has all the hallmarks of a rushed piece of legislation. It takes no account of the present technology available which allows users to hide their identity online. This is perhaps the critical point as to why the legislation will fail at reducing piracy. Those who carry out piracy on a large scale will not be deterred by the frail and misdirected measures. The legislation also relies on rights holders utilising flawed measures to attempt to identify infringers which can lead to misidentification. The doctrine of vicarious liability for network owners puts an unnecessary restriction on the digital infrastructure which will jeopardise the vision set out in the Digital Britain report. This vision is further jeopardised by the short-sighted inclusion of the technical measures, which could see households disconnected from the internet. In the digital age internet

[63] Directive 2009/140/EC of the European Parliament and of the Council of 25 November 2009 amending Directives 2002/21/EC on a common regulatory framework for electronic communications networks and services, 2002/19/EC on access to, and interconnection of, electronic communications networks and associated facilities, and 2002/20/EC on the authorisation of electronic communications networks and services OJ L 337/37.
[64] Matt Warman, 'Council tenants in 'digital apartheid' (The Telegraph, 6 December 2011) <http://www.telegraph.co.uk/technology/broadband/8938499/Council-tenants-in-digital-apartheid.html> accessed 30 April 2012.
[65] Art. 10 (2) European Convention on Human Rights and Fundamental Freedoms.
[66] Bonadio (n 27) 624.
[67] ibid.

disconnection has serious repercussions for freedom of expression and as such a judicial ruling must be a prerequisite – something that the DEA does not expressly allow for. In this way the DEA is short-sighted, rash and counter-intuitive to the more sensible ways of dealing with piracy.

4 REVIEW OF THE DIGITAL ECONOMY ACT

4.1 Introduction

After the previous section's examination of the DEA, the wider implications and reactions to the legislation must now be discussed. Unsurprisingly, the DEA was subject to hostile reaction from ISPs, who were dismayed by the burdensome obligations the legislation put upon them. This resulted in a judicial review of the legislation in the High Court brought by BT and Talk Talk, the results of which shall be examined in this section.[68] The DEA was also mentioned in the review, *Digital Opportunity - A review of Intellectual Property and Growth*, which was commissioned by David Cameron to evaluate how well the current IP framework supports economic growth and innovation.[69] This section will therefore also briefly discuss the findings of the Hargreaves Review and what implications they may have for the future of the DEA.

4.2 Judicial Review

It must first be established that many of the grounds of review are for breaches of European Union (EU) law, which the courts are bound to apply. The first ground for review was the submission that the initial obligations were in breach of the Technical Standards Directive (TSD).[70] Article 8(1) TSD requires Member States to inform the European Commission (Commission) of 'any draft technical regulation' along with a statement explaining why such a regulation is necessary. Technical regulations are compulsory 'requirements or rules on services' for the provision of a service in a Member State.[71] The purpose of the TSD is to allow to the Commission to propose changes to reduce any restraint on free movement in the internal market caused by the draft technical regulation. The DEA draft initial obligations code was not communicated to the Commission. Parker J dodged the issue and held that because the initial obligations are not yet legally enforceable against any individual, they do not satisfy the requirement of 'legal effect' for the TSD to apply under the case law of the Court of Justice (CoJ).[72] However, he effectively conceded that once the code is enacted, the initial obligations will fall under the TSD.[73] It is unclear whether the Commission will find the initial obligations restrictive of free movement and if they will find them necessary.

---

[68] *The Queen (on the Application of), British Telecommunications Plc, Talktalk Telecom Group Plc v The Secretary of State for Business, Innovation and Skills* [2011] EWHC 1021.
[69] Ian Hargreaves, 'Digital Opportunity - A Review of Intellectual Property and Growth' (2011) <http://www.ipo.gov.uk/ipreview-finalreport.pdf> accessed 30 April 2012.
[70] Directive 98/34/EC of the European Parliament and of the Council of 22 June 1998 laying down a procedure for the provision of information in the field of technical standards and regulations L 204/37.
[71] ibid Art. 1(11).
[72] *British Telecommunications Plc* (n 1) 19.
[73] ibid.

The second ground for review was for a breach of the 'mere conduit' provisions of the e-Commerce Directive (ECD).[74] Under these provisions ISPs cannot be held liable for the information transmitted through their service. Parker J held that under the initial obligations ISPs incurred no liability, as they are only required to notify subscribers and compile CILs.[75] He argued that the ECD was carefully drafted so as to strike a fine balance between competing interests (ISPs and rights holders), and to conclude that ISPs were liable under the initial obligations would upset this balance.[76] However, it seems erroneous to conclude that ISPs are not liable under the DEA when s.14 presents a £250,000 fine for non-compliance.

The third main ground for review was for a breach of the Authorisation Directive (AD).[77] ISPs argued that DEA infringes on the directive because it discriminates against larger ISPs because smaller ones and mobile network operators are not subject to the provisions of the Act. Parker J rejected this, holding it is reasonable and proportional to target large ISPs first. However, Parker J ruled that ISPs should not be required to bear any part of Ofcom's or the appeals body's costs of setting up, monitoring and enforcing the scheme: these would amount to administrative charges on ISPs extending beyond the exhaustive description of recoverable administrative charges under Art. 12 AD. As mentioned in Section 3, the financial burdens set to be placed on ISPs are onerous, and so any limitation of this burden is to be welcomed.

The fourth prominent ground of review was under the very wide heading of 'proportionality'. The need for proportionality comes from a general principle upheld by both European Court of Human Rights (ECtHR) and CoJ jurisprudence. So, in the case of the DEA, rights such as privacy of communications and freedom of expression (under Art.8 and Art.10 of the ECHR, and Art.7 and Art.11 of the Charter of Fundamental Rights) can only be restricted if done so by a law which is proportionate and necessary. The restriction by law must be pursuant to legitimate aim and must only be so far as is necessary to pursue that aim. Parker J held that this sort of balancing act should be left to parliament, who should be given a wide margin of discretion when it comes to balancing two or more fundamental rights.[78] It is submitted here, however, that it should be for the courts to ultimately strike a correct balance between rights, as held by the French Constitutional Court in the review of France's own copyright enforcement legislation.[79] Furthermore, Parker J was reluctant to find the legislation disproportionate because of what he believes was a 'lengthy process of consultation' which provided parliament with sufficient insight to correctly balance the interests at

[74]Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (Directive on electronic commerce) L 178/1, Art. 12, 15.
[75] *British Telecommunications Plc* (n 1) 27.
[76] ibid 23.
[77] Directive 2002/20/EC of the European Parliament and of the Council of 7 March 2002 on the authorisation of electronic communications networks and services (Authorisation Directive) L 108/21.
[78] *British Telecommunications Plc* (n 1) 46.
[79] Enrico Bonadio, 'File-sharing, copyright and freedom of speech' (2010) 33(10) EIPR 619, 624.

stake.[80] This argument is unconvincing considering that it has since been revealed the government put undue weight on unchallenged evidence provided by rights holders when drafting the legislation.[81]

Parker J also put a gloss on the argument that the DEA will stifle access to the internet. He reasons that it is premature to conclude whether there will be any chilling effects, and if there are, these are proportionate to enhancing copyright protection.[82] This leads to perhaps the pertinent reason as to why the judicial review failed on the proportionality ground: counsel for the ISPs did not present a less intrusive alternative system for protecting copyright online. It is submitted that since the judge was not presented with an alternative which proved the measures in the DEA were unnecessary, he could not rule that the legislation was disproportionate. However, over all Parker J seemed too eager to gloss over the more serious concerns about the necessity of the measures in the DEA and gave too much weight to the view that the legislation was carefully drafted by parliament – which, as we have seen in Section 3, it was not.

Although the initial review was unsuccessful for the ISPs, it appeared in the aftermath that the appeal could be successful.[83] Unfortunately, it has recently been announced that the Court of Appeal has dismissed the ISPs' challenge.[84]

4.3 The Hargreaves Review

In 2010 Professor Ian Hargreaves was commissioned by David Cameron to chair a review of the IP framework and how it supports economic growth. The result was the report *Digital Opportunity - A review of Intellectual Property and Growth* which was published in May 2011. The review's first recommendation is relevant to the DEA. It holds that IP policy should be 'driven as far as possible by objective evidence'.[85] Although online piracy is often claimed to be the ultimate bane on creative industries, there is little reliable data to support this assertion. The review 'failed to find a single UK survey that is demonstrably statically robust' and much of the data cannot be reconciled, as mentioned in Section 2.[86] This is problematic given that the DEA came into being because of studies like the BPI report, the methodology of which is not available for peer review. The government itself has admitted as much.[87] It is worrying that a piece of legislation which has completely altered the UK's digital landscape, with

[80] ibid.
[81] Editorial, 'Government admits Digital Economy Act based solely on rights holders' data' (*V3*, 22 November 11) <http://www.v3.co.uk/v3-uk/the-frontline-blog/2126931/government-admits-digital-economy-act-solely-rights-holders> accessed 30 April 2012.
[82] *British Telecommunications Plc* (n 1) 52.
[83] Editorial, 'Appeal over legality of Digital Economy Act could be successful, judge says' (*Out-Law.com*, 31 October 2011) <http://www.out-law.com/en/articles/2011/october/appeal-over-legality-of-digital-economy-act-could-be-successful-judge-says/> accessed 30 April 2012.
[84] Editorial, 'Digital Economy Act not in breach of EU laws, Court of Appeal rules' (*Out-Law.com*, 6 March 2012) <http://www.out-law.com/en/articles/2012/march1/digital-economy-act-not-in-breach-of-eu-laws-court-of-appeal-rules/> accessed 30 April 2012.
[85] Hargreaves (n 2) 8.
[86] ibid 69.
[87] Editorial (n 14).

potentially disastrous effects for internet subscribers, seems to have been based on unreliable evidence.

The Hargreaves Review does not itself undertake an analysis of the DEA. In fact, the DEA is not mentioned very often in the review. Perhaps, not being a lawyer, Professor Hargreaves felt under-qualified to critically evaluate the DEA. However, the review recommends that the implementation of the DEA be carefully monitored so that 'the approach can be adjusted in the light of evidence'.[88] Given that Professor Hargreaves may well recognise that the creation of the DEA was not based on much evidence, he felt it necessary to recommend that the government should respond to evidence when enforcing the DEA. Therefore, for example, a large number of appeals from wrongfully identified subscribers could indicate to the government that the initial obligations are not working as intended and should be scrapped. It is also important that the response to the CIRs be carefully monitored. This is an opportunity to document how educating consumers on copyright law affects their willingness to infringe, and to determine if the threat of legal action or internet disconnection is really necessary.[89] Even if the DEA is to fail as a proportionate way of tackling online piracy, data should be gathered so policy makers can avoid making similar mistakes in the future.

The review is also critical of the conditions in which the bill was passed. The Hargreaves Review suggests that because of the imminent general election and the need to get legislation passed, the government could not withstand pressures to amend the bill.[90] Lord Puttman admitted that the degree of lobbying on behalf of rights holders was 'quite destructive'.[91] This view does not sit well with Parker J's conclusion that the DEA was a result of considerable consultation which allowed the government to successfully balance the interests at stake. The fact that the two leading ISPs, whose cooperation is surely essential, are challenged the legality of the DEA shows the lack of business consensus.[92] . It is submitted that the Hargreaves Review is supportive evidence that the DEA was rushed and is poorly drafted as a result.

4.4 Conclusion

The ISPs recently failed in their judicial review of the DEA and therefore it appears that it will be implemented in the future. The decision does little to redeem the DEA, as the continuing adverse reaction of ISPs is indicative of their lack of respect for the legislation. There are alternative ways to tackle piracy, with which ISPs would gladly cooperate, and so the DEA is unnecessary. These will be explored in the following section. It is perhaps regrettable that the Hargreaves Review did not take a more critical approach towards the DEA. It illuminated some of the problems with the passing of the bill, such as disputed evidence relied upon in consultation and the influential lobbying from interested parties. It failed however to explore the DEA's probable negative effect

[88] Hargreaves (n 2) 6.
[89] ibid 81.
[90] ibid 92.
[91] ibid 93.
[92] ibid.

upon the UK's digital infrastructure and economy (as described in Section 3), which was the overall objective of the review.

5 SOLUTIONS

5.1 Introduction

There may be doubts as to the economic damage file-sharing causes the entertainment industries, but there is no doubt that file-sharing is happening on a large scale.[93] It is so common place in the digital society, that many people, especially generation Z, do not recognise it as inherently wrong.[94] It is possible to say the file-sharing, in and of itself is a socially desirable expression that should not be limited. File-sharing may even increase the number of new works created by providing a huge commons as described in Section 2. Therefore, any policies attempting to reduce piracy or to secure better reward for copyright owners should be proportionate and sensible, and not damage the benefits to society that file-sharing brings. Legislation such as the DEA is an example of this. It is intrusive, expensive and inefficient. There are alternatives. This section will examine possible solutions to file-sharing, from maintaining the status quo with consumer-focused innovation to the possible introduction of a non-commercial use levy.

5.2 Underlying considerations

Despite the Hargreaves Review's rather timid treatment of the DEA, the report makes a number of obvious, but important recommendations. Firstly, as mentioned in the previous section, there is the broad recommendation that the Government 'should ensure that development of the IP System is driven as far as possible by objective evidence'.[95] The review also makes it clear that the argument of securing 'just reward' for creators in favour of extending or strengthening copyright law is no longer satisfactory.[96] Creative industries are an important sector in the UK economy and therefore attention must be paid to hard economic evidence to determine how a policy will promote the creation of new works – the underlying aim of copyright law. Legislation such as the DEA does not meet these criteria. The coalition government agreed with this recommendation in the official response, stating that it 'will in future give limited weight in IP policymaking to evidence that is not sufficiently open and transparent in its approach and methodology'.[97] This is to be welcomed and it is hoped that the government stand by this statement. Objective economic evidence must be firm in the mind of any proposals to address digital piracy.

---

[93] Ian Hargreaves, 'Digital Opportunity – A Review of Intellectual Property and Growth' (2011) 69.
[94] Lior Strahilevitz, 'Charismatic Code, Social Norms, and the Emergence of Cooperation on the File-Swapping Networks' (2003) 89 Val L Rev 505, 542–44.
[95] Hargreaves (n 1) 30.
[96] ibid 19.
[97] HM Government, 'The Government Response to the Hargreaves Review of Intellectual Property and Growth' (2011) 3 <http://www.ipo.gov.uk/ipresponse-full.pdf> accessed 30 April 2012.

The second wide recommendation is that IP framework should be adaptive.[98] This is in support of the evidence recommendation, which would see the Intellectual Property Office (IPO) given statutory powers to publish reports and recommendations with the aim of 'ensuring that the UK's IP system promotes innovation and growth through efficient, contestable markets.'[99] This monitoring of the IP framework is important in an age of rapid technological development. Policy must have regard to new technologies if we are to avoid a repeat of the current situation where rights holders are using copyright law to protect as a shield to innovation.

5.3 Maintaining the status quo

It is submitted illegal file-sharing could be tackled with very little change to the current copyright framework. There is no method that will completely and totally prevent copyright infringement, and rights holders recognise this. In the physical world, copyright infringement takes place, but it is on a level at which rights holders can turn a blind eye to it.[100] It is impossible to police the internet effectively, and so rights holders should not invest in piracy prevention. Moreover, as was discussed in Section 2, a pirated copy does not necessarily represent a lost sale, so levels of piracy do not adequately represent economic loss for rights holders. Changes should be made to increase legitimate content delivery methods to maximise profits for rights holders, but a certain amount of piracy will always exist, as it does in the physical world. The law could introduce several measures to make it easier for rights holders to operate in the digital market.

A change in attitude by the entrenched industries is necessary. They should abandon the use of copyright law as a shield to innovation. It is important for these industries to discover why consumers pirate their works. Firstly and most obviously, the works are available for little or no cost. But this is not the only reason why works are downloaded illegally, because some consumers pay for pirated content. They pay subscription fees for file-hosting websites (to achieve the fastest download rates) or donate to keep private P2P networks running.[101] Take video content for example. The consumer in the digital age favours pirated content because they can access the content on demand. They do not have to endure an aggregate of fifteen minutes of advertisements for every hour of content. They do not have wait months until it is released on DVD and bother with the hassle of physical copies which can be lost or damaged. They can watch television which may not even be available in their country – file-sharing transcends all licencing barriers. They can download content in high-definition, which may not be available to them through legitimate means. They can watch the file on any device, from their PC to their television to their smart phone – pirated copies are not restricted by DRM. This

[98] Hargreaves (n 1) 96.
[99] ibid.
[100] Colin Nasir, 'From scare tactics to surcharges and other ideas: potential solutions to peer to peer copyright infringement: Part 3' (2005) 16(5) Ent LR 105.
[101] Editorial, 'Premium-rate calls watchdog to join battle against pirates' (*The Register*, November 2011)
<http://www.theregister.co.uk/2011/11/03/prs_copyright/> accessed 30 April 2012.

is what the consumer in the digital age demands. It will be very difficult for the entrenched entertainment industries to combat piracy unless legitimate content appeals to consumer demand. Creative industries must engage with the consumer instead of attempting to punish them for acquiring their products.

The industries are slowly realising this. In the same way that they initially hailed the VCR as the bringer of the apocalypse for the film and television industry, only to later embrace it as a way to sell huge amounts of catalogued material and profit from it.[102] Content delivery models, similar or inspired by P2P are becoming more prominent as a legitimate means of content delivery. The range of innovative ways in much music, video and video games are delivered to consumers is testament to the changing face of content delivery. Apple's iTunes has been a success for the music industry.[103] In this model a huge selection of individual songs are sold at relatively cheap price and then downloaded to the customer's hard drive. Apple recently removed the DRM from its music to allow songs to be transferred and played on any device.[104] Presumably this was in response to consumer demand. This is an example of control over the work being rescinded in order to attract the modern digital consumer. Advertisements can also substitute income from music sales. The hugely popular Spotify offers free music streaming, with short advertisements in between every few tracks, with the opportunity to upgrade to a premium subscription service which removes the advertisements. For video, the UK market has lagged behind the US, but we have movie and television streaming services, such as the recently launched Netflix. For a price of £5.99 per month, the customer has access to a vast library of television and films which can be streamed, without limit, in high definition to PCs and televisions. In April 2011, Netflix announced it had 23.6 million subscribers in the US.[105] A combination of large selection and freedom of use, which make pirate copies attractive, is popular with consumers.

Some video games developers are embracing the free-to-play model as mentioned in Section 2. This includes, for example, even AAA video games such as Valve's Team Fortress 2. The multiplayer game is available to download and play free of charge. Once in game, the player can continue to pay nothing or pay small amounts for different in-game weapons and vanity items through a microtransaction model.[106] The switch from retail only to free-to-play has been hugely successful for Valve, which saw revenues

[102] Colin Nasir, 'Taming the beast of file-sharing - legal and technological solutions to the problem of copyright infringement over the Internet: Part 1' (2005) 16(5) Ent LR 50, 5.
[103] Editorial, 'ITunes' success revolutionizes music business' (*Leader Telegram*, 9 October 2011) <http://www.leadertelegram.com/news/daily_updates/article_90223553-7f46-561e-b864-1f18736824ba.html> accessed 30 April 2012.
[104] Michael, 'Apple announces DRM Free song in iTunes with Higher Quality Audio' (*Apple Gazette*, 2 April 2007) <http://www.applegazette.com/itunes/apple-announces-drm-free-song-in-itunes-with-higher-quality-audio/> accessed 30 April 2012.
[105] Erick Schonfeld, 'Netflix Q1 Earnings Up 88%, Adds 3.M Subscribers' (*Seeking Alpha*, 25 April 2011) <http://seekingalpha.com/article/265310-netflix-q1-earnings-up-88-adds-3-m-subscribers> accessed 30 April 2012.
[106] Wesley Yin-Poole, 'Valve: Team Fortress 2 F2P switch a resounding success' (*EuroGamer.net*, 8 March 2012) <http://www.eurogamer.net/articles/2012-03-08-valve-team-fortress-2-f2p-switch-a-resounding-success> accessed 30 April 2012.

from the game increase by a factor of twelve.[107] The free-to-play model can also be supported by in-game advertising. These examples above demonstrate how innovation in content delivery is underway. Those who take the risk with these new systems will most likely reap the benefit. In truth, file-sharing has spurred this innovation and is forcing companies to meet the needs of the consumer, which in itself is advancing consumer welfare.

There is a way in which new methods of content delivery can be aided and promoted. The government can attempt to reduce the arduous burden of acquiring licences. For example, the UK's first video streaming service, BBC iPlayer, was a result of *five* years of licence acquisition.[108] This inefficiency needs to be improved to provide healthy and contestable markets for new content systems which will attract consumers away from piracy. A way to this is through a Digital Copyright Exchange, perhaps the best proposal in the Hargreaves Review. A Digital Copyright Exchange would exist as a network of internet accessible databases which would, amongst other services, record the ownership of all rights attached to the work and provide automated licencing via standard terms.[109] This would have the effect of greatly reducing transaction costs, time taken to secure licencing, and provide certainty for proposed business models reliant on licencing. This would result in more competition and innovation in the content delivery market, from which both creators and consumers would benefit. More legitimate content delivery systems would reduce the levels of illegal file-sharing.

There are still some questions which need to be answered surrounding a Digital Copyright Exchange, such as, who will build it, who will bear the costs and how will rights holders be incentivised to put their work on it.[110] To compliment a Digital Copyright Exchange, '[t]he UK should support moves by the European Commission to establish a framework for cross border copyright licensing'.[111] Despite the EU's devotion to the single internal market, there is currently no method of securing licences for more than one EU country at a time, which poses disproportionate transaction costs on those seeking to establish themselves in an EU-wide market. If this were to change, a Digital Copyright Exchange could be used to licence UK works across the EU.

Assuming draconian anti-piracy measures are abandoned, innovation can tackle piracy while at the same time preserving the beneficial aspects of file-sharing, self-expression, individual autonomy and creative collaboration. Investment in new content delivery methods also increases competition and consumer welfare, and can increase profits for rights holders.

[107] ibid.
[108] Hargreaves (n 1) 29.
[109] ibid 31.
[110] ibid 33.
[111] ibid 37.

5.4 A non-commercial use levy

A more radical solution to illegal file-sharing is the possible introduction of a non-commercial use levy, as proposed by Neil Netanel.[112] This would be a levy imposed on the sale of all products which can be used for, or can contribute in some way to file-sharing. The bounty from this levy would then be redistributed to rights holders. In return, non-commercial file-sharing for private use would be legalised. How this will work in practice is detailed below.

With the introduction of a non-commercial use levy, file-sharing of most copyright protected works would be legal, including works of entertainment such as music and video. The exception would also apply to the streaming of copyright protected works such as on YouTube.[113] Only file-sharing for non-commercial purposes would be legal; those seeking to profit in any way would still need to acquire a licence to use the work. Since non-commercial file-sharing would not constitute copyright infringement, providers and facilitators of, for example, P2P networks would not be liable under s.16(2) CDPA for 'authorisation'.

This has a number of positive effects. The important cultural benefits of file-sharing would be preserved. Rights holders will save money by avoiding costly and futile litigation against individual file-sharers or individual facilitators. Rights holders, particularly the recording industry and the film industry, could rebuild their reputation with consumers which has been tarnished by years of aggressive litigation. A non-commercial use levy which legalised file-sharing would also bring the law into line with the public opinion that 'there is nothing morally wrong with file-sharing'.[114] This is a major step forward, as digital copyright law has for too long been completely at odds with public perception.

As discussed in Section 2, the current copyright model is a proprietary model. Copyright grants authors an exclusive right to exploit their property which is a result of their labour. A non-commercial use levy would be contrary to this theory, but it would remedy some of the problems with the proprietary model. Arguably, the proprietary model is unfit for the digital age and it exists as an anachronism. In the digital arena, there are no physical copies to reproduce, reproduction can be done instantly and with no loss of quality – the rights holder incurs no cost from reproduction; and with file-sharing all costs are borne by the consumer (PCs, internet access, media devices).[115] The current copyright framework does not take this into account. The battle against file-sharing is arguably an attempt by dominant companies to extend the copyright monopoly that they have had over hard copies.[116] The non-commercial use levy would

[112] Neil Netanel, 'Impose a Noncommercial Use Levy to Allow Free Peer-to-Peer File Sharing' (2003) The University of Texas School of Law: Law and Economics Working Paper No. 009.
[113] ibid 37.
[114] ibid 18.
[115] ibid 28.
[116] ibid 27.

embrace the free reproduction costs, tax the costs borne by the consumer for file-sharing technology, and use the proceeds to compensate rights holders.

It would also have the effect of remedying a unique problem in the economics of IP law. This has been labelled as 'the tragedy of the anticommons'. In the anticommons, multiple owners have the right to exclude but no one has the privilege to use.[117] Copyright law confers many exclusive rights on authors, which are often assigned to various others, creating a complex web of overlapping rights. This results in significant transaction costs for the use of copyright protected works, which means they will be used less. This is tragedy which economists are concerned with – underuse – and results in a diminutive public domain.[118] It can be argued that the purpose of copyright law, which was championed by the proprietary model, is no longer to protect authors. It is to protect authorship. The distinction is crucial.[119] Material in the public domain can be used by any member of the public, without incurring costs. This makes creation of new works easier in practice and more economically viable, and as such promotes creativity and valuable social interaction. A non-commercial use levy which legalises file-sharing would essentially bolster the public domain and encourage use of works, avoiding the tragedy of the anticommons. As such, the non-commercial use levy would expressly permit the creation of non-commercial derivative works, so long as the used work was credited.[120] It could achieve this and still provide ample compensation to rights holders.

The non-commercial use levy would be applied to any product or service 'whose value is substantially enhanced by P2P file-sharing'.[121] This would apply to blank media, such as writable CDs and DVDs, media devices such as MP3/MP4 players, media servers, external hard drives, PCs and laptops, commercial suppliers of P2P software and ISPs. The responsibility for implementing the non-commercial use levy should be given to a government body, for example, a newly created division of the Intellectual Property Office. This body would decide the rate of the levy for each product, and which products and services the levy will apply to. The levy will be negotiable and industries affected will be able to debate the levy rate before the body, allowing evidence to be taken into account. The ability of the levy to change frequently also ensures it can react to rapid technological developments, as recommended by the Hargreaves Review. The suppliers of goods and services subject to the levy would also benefit from the legalisation of file-sharing because the levy, by definition, only applies to those products and services whose value is enhanced by file-sharing. They would also save money from abandoning the need to implement DRM-compliant technology on their devices. In fact, these industries could invest in enhancing their relationship with file-sharing technology to the benefit of consumers and the public in general.[122]

---

[117] Samantha Leung, 'The commons and anticommons in intellectual property' (2010) 16 UCL Juris Rev 16, 23.
[118] ibid.
[119] Jessica Litman, 'The Public Domain' (1990) 39 Emory LJ 965, 969.
[120] Netanel (n 16) 39.
[121] ibid 43.
[122] ibid 21.

This thesis cannot go into detail as to the economics of the calculation of the levy, but it should seek to 'yield the purported benefits of proprietary copyright without imposing its costs'.[123]As has been determined, viewing every illegal downloaded copy of a work as a lost sale is erroneous, but this, in principle, can be used to calculate the levy. However, there is no reason that calculating the levy should take into account the price of physical copies – it will take into account only the work itself, and not packaging, distribution and other physical costs. Of course, the calculation of the levy would be subject to rigorous and ever-changing economic analysis, but Netanel provides a rough but illuminating calculation for the implementation of the levy in the US.[124] For example, if all the relevant industries have gross annual revenue of $38.2 billion, and half of this is attributable to the 'physical' aspect of sales, then the net revenue for the works is $19.1 bn. The percentage of 'anticipated' loss due to file-sharing varies depending on the industry (music being the highest; an estimated 25%) but the total estimated 'loss' is $2.51 billion – which the levy must yield. Then total sales for all the P2P-enhanced products and services are taken are calculated. The average levy rated on this figure needed to generate the amount 'lost' due to file-sharing is a meagre 4.06%.[125] This is a small price to pay for enhanced social and cultural interaction, freedom of access and use of a whole domain of works.

Perhaps the most challenging aspect of implementing a non-commercial use levy is the redistribution of the funds. The proceeds should be distributed to rights holders in proportion to the number of downloads, views or listens (for streams) and uses.[126] The technology now exists to do this, and any costs can be cut by use of effective statistical sampling.[127] Under the levy, derivative works would credit the underlying work, allowing rights holders to reap the rewards of adaptions. Derivative works would need to retain 'copyright management information' of the underlying work, but this would not be difficult to enforce – P2P networks could easily apply this as a rule if they want to enjoy the exception. Although the redistribution of the levy's bounty may be imprecise initially, the body responsible for implementing the levy will have the ability to adapt to changing information gathering technologies and to use the evidence before them to make any changes necessary to redistribution.

The main obstacle to the implementation of a non-commercial use levy in the UK is the single market. This is symptomatic of a wider problem of designing national digital copyright policy to deal with copyright infringement which knows no national barriers and operates on a global scale. If a levy is introduced to increase the prices of goods and services in the UK, then foreign competitors may gain an unfair advantage. For example, Rue Du Commerce, a French online retailer, was recently awarded damages of €1m with interest from Copie France (responsible for implementing the private

[123] ibid 37.
[124] ibid 63-65.
[125] ibid 65.
[126] ibid 53.
[127] ibid 55.

copying levy) for loss of sales.[128] It is argued that the levy is fundamentally an excellent proposal, which preserves and even encourages the valuable social and cultural interaction of file-sharing while at the same time rewarding creators. However, the implementation of the levy in the UK would require significant research and discussion, namely due to the problems with competition law and the free movement of goods and services.

5.5 Conclusion

There are alternatives to draconian legislation such as the DEA. Even strategy as simple as renewed focus on innovation and communication with consumers could do a great deal to reduce the negative effects of piracy. Those who are innovating are already reaping the benefits of new content delivery models. If a more grounded approach is necessary, then a non-commercial use levy appears to have the potential to appeal to both consumers and industries alike. Further research is needed to determine precisely how it would operate in the UK.

6 CONCLUSION

This thesis has hopefully presented three main arguments which should inform future developments of digital copyright law. The first, presented in Section 2, is that illegal file-sharing is not inherently adverse to the overarching objective of copyright law – to encourage the creation of new works. New works require others for inspiration, which is achieved by the large and active commons which file-sharing networks provide. File-sharing is important for social and cultural communication and is protected by Art. 10 ECHR. File-sharing networks also provide the opportunity for others to collaborate in the creation of new works, explore derivative works which promote identity and self-realisation. The technology behind file-sharing has also spurned demand for new content delivery methods and payment models. The increased competition file-sharing technologies and traditional delivery methods have led to a host of new services such as video and music streaming applications and free-to-play video games. This has undoubtedly increased consumer welfare and will continue to increase if entertainment industries engage with consumers who practice file-sharing. It is submitted that file-sharing *per se* is beneficial to society.

The second argument presented in Section 3 is that legislation which seeks to use harsh measures against illegal file-sharing can never be successful. The DEA is presented as an example. The legislation cannot hope to contend with technological developments in the digital arena which allow file-sharers to easily mask their identity to avoid detection under the DEA, which could also lead to misidentification of innocent users. The DEA also has the effect of putting a burden of providers of services such as internet, Wi-Fi and 3G. Services providers will be obligated to notify infringing subscribers – which may be an onerous burden for some providers. Eventually, providers may be

[128] Editorial, 'One million euros in damages and interest for Rue du Commerce' (*E-Commerce Facts*, 8 December 2011) <http://www.e-commercefacts.com/news/2011/12/rueducommerce-copie-franc/index.xml> accessed 30 April 2012.

forced to terminate or throttle connections of repeat offenders, even though internet access is regarded as a fundamental right. In a time when the UK is recovering from the worst financial crisis in decades, as we seek to improve our infrastructure to allow our creative industries to compete in the global digital market, it is unforgivable that the government should seek to put such significant burdens on providers and users alike.

Although the judicial review of the act has proven unsuccessful, it has recently been announced that the DEA will not be implemented until 2014 at the earliest, with the Internet Service Provider's Association stating this gives hope for rational alternative measures to be explored.[129] It is submitted that legislation as harsh as this does harm to almost all parties involved, to users, service providers, to the economy, and even to rights holders as they bear costs of detection. The DEA is unnecessary for achieving the aims of copyright law set out in Section 2.

The third proposition, presented in Section 5, is that there are methods of assisting rights holders in dealing with piracy which do not require draconian measures. Right holders are already benefiting from innovation which seeks to appeal to consumers. Services such as Spotify, iTunes and Netflix are hugely popular. It is submitted that the piracy-boom in the last decade was a result of the entrenched industries disregarding new technology, and only now are these industries meeting consumer demand. The law can make it easier for these industries to create innovative content delivery systems. For example, a Digital Copyright Exchange as recommended by the Hargreaves Review would make licencing easier and would reduce transaction costs, making new content delivery services more economically attractive. These measures would increase the amount of legitimately delivered content, which should reduce the amount of piracy while at the same time preserving the availability and the ease of access to creative works.

Another, more radical alternative method of aiding rights holders is to introduce a non-commercial use levy. This would legalise all non-commercial file-sharing which would be of massive social and cultural benefit, while at the same time providing remuneration to rights holders. This, *prima facie*, is a very attractive solution but in order for it to be implemented in the UK further consultation is needed. It requires a fundamental change in copyright law – the departure from the traditional proprietary model. A body would also need to be set up to calculate, oversee and distribute the levy. Moreover, there may be competition problems within the UK and the EU, as foreign products and services could have an unfair advantage as they would most likely not be subject to the levy. A non-commercial use levy provides a perfect solution in principle, but further research is needed to discover if it could work in practice.

In conclusion it is submitted that file-sharing can co-exist with the current copyright regime, without the need for additional enforcement legislation such as the DEA. File-

[129] Editorial, 'Digital Economy Act's anti-piracy measures are delayed' (*BBC News*, 26 April 2012) <http://www.bbc.co.uk/news/technology-17853518> accessed 30 April 2012.

sharing has become such a fundamental part of digital culture that there should be no attempt to stifle it, nor would it be possible to do so. It should be for the entrenched industries to combat piracy by offering consumers viable alternatives instead of lobbying government to use the law against consumers. If there is to be new legislation introduced, it should be something similar to a non-commercial use levy which can compensate rights holders while preserving humanity's cultural heritage in an open, diverse, and free commons.

JUDICIAL REVIEW UNDER THE HUMAN RIGHTS ACT: A CULTURE OF JUSTIFICATION

*THOMAS RAINE**

The approved function of the judiciary is consistently debated, especially since the introduction of the Human Rights Act 1998. The Act outlined that the courts were to review the actions of primary decision makers in light of convention rights. Arguably, the effect of this act has resulted in a culture of justification whereby the political branches of the government now have to justify their decision with reference to the Human Rights Act. However, this justificatory burden has been limited as the courts have discharged their role of protecting convention rights where they have to consider a case of political controversy through the concept of the discretionary area of judgement. This article prescripts the correct approach to which the court should take, outlining that a degree of appreciation for the political institutions can be made where it is justifiable for a degree of deference to be made to the political institution. This will allow the justificatory burden on decision makers to be fully established.

1 INTRODUCTION

As Dyzenhaus puts it, 'The role of judges in the legal order has always been controversial'.[1] Within the UK, that controversy has increased with the enactment of the Human Rights Act 1998 (HRA). The Act has given the judiciary the authority to review Acts of Parliament, assessing their compatibility with the European Convention on Human Rights (ECHR). Section 4 of the HRA permits judges to issue a 'declaration of incompatibility', an explicit statement of a statute's inconsistency with the Convention. Furthermore, Section 6 of the HRA requires public authorities, including the executive, to act in a manner compatible with Convention rights. The courts are charged with the responsibility of reviewing the decisions of both Parliament and the executive, the elected branches of government, to ensure they do not breach human rights. Therein lies the controversy. As Ewing argues, the HRA represents an 'unprecedented transfer of political power from the executive and legislature to the judiciary'.[2] The Act requires the courts, unelected and democratically unaccountable, to substantively review the actions and decisions of the elected branches.

How should judges review the actions and decisions of the elected branches under the HRA? It is to this question that this essay is dedicated. Firstly, the effect of the HRA upon the UK's constitutional order will be examined. It is to be argued that the HRA has cultivated a new culture at the heart of our constitution, a 'culture of justification'.[3] In this new culture, any act or decision of the elected branches that limits a Convention

* Thomas Raine, Newcastle University. Law LLB Stage Three.

[1] David Dyzenhaus, 'The Politics of Deference: Judicial Review and Democracy' in Michael Taggart (ed), *The Province of Administrative Law* (Hart 1997) 279.

[2] KD Ewing, 'The Human Rights Act and Parliamentary Democracy' (1999) 62 M.L.R. 79, 79.

[3] Etienne Mureinik, 'A Bridge to Where? Introducing the Interim Bill of Rights' (1994) 10 S.A.J.H.R. 31, 32.

right must be objectively *justifiable*; it must be supported by rational and cogent argument. The elected branches can no longer rely upon their democratic credentials or superior constitutional status to justify their actions. Central to this end is proportionality which has replaced *Wednesbury* irrationality as the standard of judicial review under the HRA. Proportionality allows for more rigorous review of primary decisions with a greater emphasis upon the reasons behind such decisions.

In light of the new culture, this essay will then consider the approach of the British courts in cases brought under the HRA. It will criticise the concept of the 'discretionary area of judgment'[4] which features heavily in the early case law under the HRA. It is to be argued that the concept has been used by the courts to avoid sufficiently robust review of decisions concerning issues of political controversy. In particular, some immigration cases will be examined to demonstrate how the discretionary area of judgment negates the requirement that any limitation of Convention rights be objectively justifiable. The cases have been chosen as in each the courts granted considerable discretion to the elected branches, simply on the basis of the subject matter of the impugned decision. Consequently the courts failed to discharge their duty of protecting Convention rights. The justificatory burden placed upon primary decision makers when limiting Convention rights was discarded.

Finally, the essay will offer a suggestion as to an appropriate approach to judicial review under the HRA that seeks to uphold the justificatory burden imposed on the elected branches, whilst ensuring that the courts are mindful of their own limitations. It is to be argued that the primary concern of the courts is proportionality. However, in applying proportionality, it is sometimes appropriate for the courts to display a degree of deference to the elected branches. The extent of judicial deference is not predetermined by the subject matter of the impugned decision. Instead the guiding principle is justifiability. The courts must ask whether it is justifiable to show deference to the primary decision maker. It will be justifiable where the primary decision maker enjoys an institutional advantage rendering it better placed than the courts to reach a reasoned decision. However it is not justifiable for the courts to defer on the basis of democratic concerns.

2 A NEW CONSTITUTIONAL ORDER

Before embarking upon an analysis of judicial review under the HRA, or prescribing an acceptable approach to such review, it is first necessary to consider the effect of the HRA upon our traditional constitutional order. In particular it must be asked: what environment are both the judiciary and the elected branches of government now operating in as a result of the Act? It is to this important question that this chapter is dedicated. It is to be argued that the HRA has introduced a new constitutional order in which any limitation of Convention rights must be objectively justifiable. Judges have been charged with the responsibility of assessing such justifications, but must also be mindful of their own institutional limitations. The HRA does not elevate the judiciary

[4] Anthony Lester and David Pannick, *Human Rights Law and Practice* (Butterworths 1999) 73.

to the position of primary decision maker. However, it does impose a justificatory burden upon primary decision makers that has not traditionally been a feature of English administrative law.

2.1 A Culture of Justification

Firstly, it is important to examine the concept of a culture of justification. It appears frequently in the academic literature relating to the HRA, particularly from those advocating more stringent judicial review.[5] However, it is not entirely clear what those invoking the concept mean when they argue that the Act has fostered a culture of justification. As it is to form an integral part of the argument it is necessary to dedicate some time to an analysis of the concept and an identification of its key features. In particular it must be asked what is meant by the term 'justification' within the context of judicial review.

A prominent advocate of the notion of a culture of justification was Etienne Mureinik, the South African human rights lawyer. Following the end of apartheid he wrote of his country's new Constitution:

> It must lead to a culture of justification – a culture in which every exercise of power is expected to be justified; in which the leadership given by government rests on the cogency of the case offered in defence of its decisions, not the fear inspired by the force of its command.[6]

This provides the founding premise of the culture of justification. Any exercise of power must be supported with sound reason and not appeals to authority. Mureinik's vision was of a South Africa free from the 'culture of authority'[7] that had sustained the apartheid regime. The culture stemmed from Diceyan parliamentary sovereignty under which the legislature enjoyed unlimited law making power. Such sovereignty had fostered an 'ethic of obedience'[8], a system of government based upon power and coercion. Through it, he argued, 'The leadership of the ruling party commanded Parliament, Parliament commanded its bureaucracy, the bureaucrats commanded the people.'[9] It was a hierarchical system of command with limited scope for insubordination or questioning. The limited nature of the judicial role was laid bare as the Nationalist Party passed discriminatory statute after discriminatory statute.[10]

---

[5] e.g. Richard Edwards, 'Judicial Deference under the Human Rights Act' (2002) 65 MLR 859; Murray Hunt, 'Sovereignty's Blight: Why Contemporary Public Law Needs the Concept of 'Due Deference'' in Nicholas Bamforth and Peter Leyland (eds), *Public Law in a Multi-Layered Constitution* (Hart 2003); Jeffrey Jowell, 'Beyond the Rule of Law: Towards Constitutional Judicial Review' [2000] PL 671.

[6] Mureinik (n 4) 32.

[7] ibid.

[8] ibid.

[9] ibid.

[10] e.g. the Population Registration Act 1950 and the Group Areas Act 1950.

Departure from this damaging order was to be achieved by imposing upon those in power an obligation to justify the actions they took. Central to this end was the new South African Bill of Rights within the South African Constitution 1996 which, Mureinik argued, proclaims 'standards of justification.'[11] Significantly, the rights contained in the Bill are not absolutes that prevail in all circumstances. Each right is capable of restriction through the general limitation clause in s.36(1) which sets out the conditions upon which restrictions of a right can be justified. A restriction must be 'reasonable and justifiable in an open and democratic society'[12] and attention must be paid to the 'importance and purpose'[13] of the limitation and whether there is a less restrictive means of achieving the same end.[14] This model allows an examination of the rationale behind executive or Parliamentary action. As Mureinik put it, 'A challenge under the Bill opens up an inquiry into the justification of the decision challenged.'[15] The burden is placed upon the primary decision maker to demonstrate the necessity of any limitation of rights within the Constitution.

The Bill itself was not the only component of Mureinik's culture of justification. There was a second, less tangible, element. As Dyzenhaus points out, key features of the culture of justification were debate, criticism and challenge.[16] As well as a new legal instrument, a more inquisitive attitude was required from the courts. Rather than merely submitting to the will of the elected branches, the courts were to critically examine the justifications behind any limitation of the rights enshrined in the Constitution. It is important at this stage to introduce the distinction between justified and justifiable.[17] When reviewing the actions of the elected branches the courts ought to consider whether such actions are *justifiable*, not whether they are justified. Such a distinction preserves the secondary nature of the judicial role and ensures that the focus of the court's review is upon the reasoning offered in defence of a limitation.

If judges were to ask whether a decision is justified, they would be seeking convergence between the decision that has been made and the decision they would have made in the same circumstances. They would be assessing the primary decision by comparison with their own personal view as to the correct course of action. As Dyzenhaus argues, under this approach the reasons behind the decision are irrelevant. All that matters is 'coincidence of content rather than the relationship between reasons and content.'[18] Judges merely ask whether the decision made coincides with their own preferred decision. They do not consider the reasons behind a particular course of action. As such, judges are not engaged in review of the impugned decision. Instead, they are asked

[11] Mureinik (n 4) 33.
[12] Constitution of the Republic of South Africa 1996, s.36(1).
[13] ibid s.36(1)(b).
[14] ibid s 36(1)(e).
[15] Mureinik (n 4) 32.
[16] David Dyzenhaus, 'Law as Justification: Etienne Mureinik's Conception of Legal Culture' (1998) 14 SAJHR 11, 13.
[17] ibid 27-28.
[18] ibid.

whether they would have made the same decision were they in the position of primary decision maker.

In contrast, when considering whether a particular decision is justifiable, judges are asking 'whether the decision maker has shown it to be defensible'.[19] Under this approach the reasons behind the decision are of central importance. Judges are charged with the task of deciding whether the reasons offered constitute a plausible, but not necessarily the only, argument in favour of the decision. The courts are not asked whether they would have adopted the same course of action as the primary decision maker, but whether that course of action is objectively justifiable. As such, judges are charged with the secondary task of reviewing the primary decision and in particular the reasons offered in its defence. They are not considering whether they would have made the exact same decision.

There are, then, two key features of a culture of justification, as identified by Mureinik. Firstly there is a legal instrument protecting important rights recognised as integral to democracy. These rights, as is the case in the South African Bill of Rights, are not framed as absolutes. Instead they permit derogation in certain circumstances. This allows for an inquiry into the justifications behind a particular action. Secondly, a new approach from the courts is required, an approach that critically examines the justifiability of the actions of the elected branches. Using these two features as points of reference one must now return to the HRA. It will be argued that these two features exist in the context of the Act in the form of the ECHR and the doctrine of proportionality, which the UK courts have adopted in response to the ECHR. As such, the HRA has introduced a culture of justification, requiring any limitation of Convention rights to be justifiable.

2.2 The European Convention on Human Rights

The HRA incorporates most of the ECHR into UK law. In the language of the preamble, those rights are given "further effect" within the domestic legal order. Of particular importance is the structure of the rights set out in the Convention. As Singh points out, there are 'three kinds of rights in the ECHR'[20]. They are:

(1) absolute and unqualified rights;
(2) rights where an interference is permitted where it is 'necessary in a democratic society'; and
(3) rights where interference is permitted where some other interest outweighs the right in question.[21]

By nature the absolute rights permit no derogation. An example is the prohibition of torture in Article 3. In contrast, the rights found in the other categories are not absolute.

[19] ibid 28.
[20] Rabinder Singh, 'The Place of Human Rights in a Democratic Society' in Jeffrey Jowell and Jonathan Cooper (eds) *Understanding Human Rights Principles* (Hart 2001) 187.
[21] ibid 187.

They allow limitation in certain prescribed circumstances. Those rights in the second category, including the right to privacy in Article 8 and the right to freedom of expression in Article 10, require any interference to be 'necessary in a democratic society'.[22] The rights in the third category are those that enjoy the least protection. An example is the right to property in Article 1 of Protocol 1 which merely requires any limitation to be in the 'public interest'.[23]

The majority of rights fall within the second and third categories and are therefore qualified, albeit in different ways. As such the Convention follows what, on Dyzenhaus's interpretation, is a 'democratic model.'[24] By permitting derogation in certain circumstances the Convention allows domestic legislatures to have a say in the determination of rights. It is not just the preserve of the courts. Dyzenhaus contrasts the democratic model with the 'liberal model'[25], such as the Constitution of the United States of America, in which rights are framed as absolutes. Under such a model it is the judiciary who have sole interpretative authority. The hallmark of the liberal model, according to Dyzenhaus, is judicial supremacism.[26]

Pressing this further, there is at the heart of the Convention a desire to reconcile two competing impulses. As Griffith reminded us, we are 'both individual and social animals'.[27] As individuals we value the rights protected in the Convention but also accept that it is often necessary to restrict such rights in the interests of the wider community. This latter concern is informed by a utilitarian instinct, requiring us to act in a manner that produces the greatest good for the greatest number.[28] It also informs the democratic model and the qualified nature of most Convention rights. There is, as Feldman observes, 'a strong element of collectivism'[29] within the Convention. The rights 'are unlikely to operate in a purely liberal and individualistic way.'[30] There is broad scope, by virtue of their qualified nature, for collective concerns to inform the application of Convention rights. The rights may be limited in light of countervailing communal interests advanced by the elected branches of government.

It is this appreciation of collectivist concerns, and recognition of the role of the elected organs in the formulation of rights, that allows the Convention to act as a vehicle for the culture of justification. For it is by permitting derogation from the protected rights that an inquiry into the justifications behind Parliamentary or executive action is opened. As Dyzenhaus argues, 'a limitation provision gives the government an

[22] European Convention for the Protection of Human Rights and Fundamental Freedoms (1950) Article 8(2), Article 10(2).
[23] ibid Protocol 1, Article 1.
[24] Dyzenhaus, 'Law as Justification' (n 17) 32.
[25] ibid 32.
[26] ibid 32.
[27] J A G Griffith, 'The Political Constitution' (1979) 42 MLR 1, 3.
[28] John Stuart Mill and Jeremy Bentham, *Utilitarianism and Other Essays* (Penguin 1987).
[29] David Feldman, 'The Human Rights Act 1998 and Constitutional Principles' (1999) 19 LS 165, 174.
[30] ibid 173.

opportunity to show that the limitation is justifiable.'[31] If rights are absolute principles over which the courts exercise sole authority, there is no scope for a justificatory examination. The only question that must be answered in such circumstances is: has a right been breached? In contrast, under the ECHR where most of the rights are not absolutes, it must also be asked: why has a right been breached and are the reasons sufficient to justify the breach? It is only the democratic model, which the ECHR adopts, that will make every exercise of power 'the proper subject of the process of justification.'[32] Only the democratic model allows an assessment of the justifiability of a restriction of a right.

2.3 The Doctrine of Proportionality

The second element of Mureinik's culture of justification was a new, more demanding, approach from the courts. This approach would subject the actions of the elected branches to a more rigorous level of scrutiny and require more in terms of justification. Closer attention would be paid to the reasons behind any limitation of individual rights. The doctrine of proportionality allows for this more intense standard of judicial review in the UK and therefore contributes to the establishment of a culture of justification. It allows our courts to consider the justifiability of primary decisions to a much greater extent than was possible under the traditional grounds of judicial review. Proportionality was specifically adopted by the UK courts to provide the more intense standard of judicial review required under the ECHR. To demonstrate the greater emphasis on justifiability it is necessary to compare proportionality with the irrationality approach that has long been a feature of English administrative law.

The traditional test for irrationality is most associated with the judgment of Lord Greene MR in the *Wednesbury*[33] case and was summarised succinctly by Lord Diplock in the *GCHQ* [34] case. Lord Diplock stated that for a decision to be irrational it must be 'so outrageous in its defiance of logic or of accepted moral standards that no sensible person who had applied his mind to the question could have arrived at it.'[35] To what extent this interpretation of irrationality has actually been adopted by the judiciary is open to question. As Craig points out, if the courts really did restrict irrationality review along the lines advocated by Lord Diplock, there would be 'almost no successful challenges of this kind.'[36] However, irrationality remains a very limited basis for judicial review. As Lord Bingham MR concluded in *Smith*, 'The threshold of irrationality is a high one.'[37] Though not impossible, it is difficult for an application to succeed.

---

[31] Dyzenhaus, 'Law as Justification' (n 17) 32.
[32] ibid 36.
[33] *Associated Provincial Picture Houses Ltd. v Wednesbury Corporation* [1948] 1 KB 223 (HL) 228-230.
[34] *R v Minister for the Civil Service, ex p Council of Civil Service Unions* [1985] AC 374 (HL)
[35] ibid 410.
[36] Paul Craig, 'Unreasonableness and Proportionality in UK Law' in Evelyn Ellis (ed), *The Principle of Proportionality in the Laws of Europe* (Hart 1999) 95.
[37] *R v Ministry of Defence, ex p Smith* [1996] QB 517 (CA) 558.

The *Wednesbury* doctrine grants the primary decision maker a very wide area of discretionary judgment into which the courts will not intrude. It is a standard of review that is 'notoriously weak'[38] in its intensity. The emphasis is on judicial restraint or, as Taggart puts it, keeping 'the judges' noses out of the tent of politics.'[39] It is symptomatic of what Harlow describes as the 'classic model'[40] of judicial review, founded upon a rigid conception of the separation of powers and a highly restricted role for the judiciary. Another significant feature of *Wednesbury* irrationality is that it places the burden of proof firmly on the applicant. The applicant had to demonstrate the irrationality of a primary decision. Furthermore, the public authority whose decision was being challenged was not required to give reasons for their actions.[41] As Taggart observes, administrative law has not historically regarded a requirement of reasons 'as an essential prerequisite to the validity of decision making.'[42] Public bodies were not obliged to demonstrate the justifiability of their decisions.

In *R (Daly) v Secretary of State for the Home Department*[43], the House of Lords explicitly adopted proportionality as the standard of review in HRA cases, replacing irrationality. In doing so the court endorsed the tripartite approach to proportionality set out by Lord Clyde in *De Freitas*.[44] When determining the acceptability of a limitation of a Convention right, the court would ask whether:

> (i) the legislative objective is sufficiently important to justify limiting a fundamental right; (ii) the measures designed to meet the legislative objective are rationally connected to it; and (iii) the means used to impair the right or freedom are no more than is necessary to accomplish the objective.[45]

In his judgment in *Daly*, Lord Steyn discussed the differences between the two grounds of review, remarking that 'the intensity of review is somewhat greater under the proportionality approach.'[46] Proportionality demands a more active role from the courts. Rather than merely considering the rationality of a decision, judges would be required to assess the balance struck by the primary decision maker. Furthermore, greater attention would have to be paid to the relative weight of rights and competing interests. In short, proportionality is a far more exacting standard of review than irrationality. As Hunt observes, proportionality requires 'a highly structured and

[38] Michael Fordham and Thomas de la Mare, 'Identifying the Principles of Proportionality' in Jeffrey Jowell and Jonathan Cooper (eds), *Understanding Human Rights Principles* (Hart 2001) 32.
[39] Michael Taggart, 'Proportionality, Deference, *Wednesbury*' [2008] NZLR 423, 429.
[40] Carol Harlow, 'A Special Relationship? American Influence on Judicial Review in England' in Ian Loveland (ed), *A Special Relationship? American Influences on Public Law in the UK* (Clarendon 1996) 83.
[41] Fordham and de la Mare (n 39) 32.
[42] Taggart (n 40) 462.
[43] [2001] UKHL 26, [2001] 2 AC 532.
[44] *De Freitas v Permanent Secretary of Ministry of Agriculture, Fisheries, Land and Housing* [1999] 1 AC 69 (PC).
[45] ibid 80.
[46] *Daly* (n 44) [27].

sophisticated analysis quite different from anything that was ever required under the more traditional grounds of judicial review.'[47]

As well as providing a more structured methodology, proportionality differs greatly from *Wednesbury* in that the burden of justifying a primary decision 'is now squarely upon the decision maker.'[48] Under *Wednesbury*, the applicant had to demonstrate the irrationality of a primary decision. In contrast, proportionality requires the decision maker to explain the necessity of any limitation of individual rights. The court must be convinced that any such limitation is justifiable within the confines of the Convention. This opens what Taggart describes as a 'justificatory gap'[49] between proportionality and *Wednesbury*. The latter requires little in terms of justification from the decision maker whereas the former imposes a strong justificatory burden. This marks a significant departure from traditional administrative law which, as noted above, had little regard for the reasons behind a particular decision.

A vivid example of the difference between irrationality and proportionality was provided in *Smith and Grady v UK*.[50] The case concerned a challenge to the prohibition on homosexuals serving in the British armed forces. The applicants alleged that an investigation into their private lives and subsequent dismissal from the armed forces on the basis of their homosexuality constituted a breach of their right to privacy under Article 8 of the ECHR. In both the Divisional Court and the Court of Appeal the applicants were unsuccessful.

The prevailing theme in the domestic courts was the limited nature of *Wednesbury*. As Leigh explains, both Simon Brown LJ and Lord Bingham MR were 'apparently inclined to find against the Crown were it not for the limitations of judicial review'[51]. Simon Brown LJ confessed that he refused the applications with 'hesitation and regret'[52] and acknowledged that in the context of the ECHR, 'the days of this policy are numbered.'[53] Whilst the Court of Appeal refused to endorse this assessment, Lord Bingham MR suggested that domestic law was unable to protect the applicants stating that it 'may be necessary' for them to pursue their claim in Strasbourg.[54] Domestic judicial review was unable to provide the rights protection that the applicants desired. Despite concerns about the policy, and its possible incompatibility with the ECHR, it could not be said

[47] Murray Hunt, 'Sovereignty's Blight: Why Contemporary Public Law Needs the Concept of 'Due Deference'' in Nicholas Bamforth and Peter Leyland (eds), *Public Law in a Multi-Layered Constitution* (Hart 2003) 337.
[48] Jeffrey Jowell, 'Beyond the Rule of Law: Towards Constitutional Judicial Review' [2000] PL 671, 680.
[49] Taggart (n 40) 439.
[50] (2000) 29 EHRR 493, see also *Lustig-Prean and Beckett v UK* (2000) 29 EHRR 548.
[51] Ian Leigh, 'Taking Rights Proportionately: Judicial Review, the Human Rights Act and Strasbourg' [2002] PL 265, 270.
[52] *Smith* (n 38) 541.
[53] ibid 542.
[54] ibid 559.

to be irrational. As Lord Bingham MR concluded, the high threshold of irrationality was not 'crossed' in this case.[55]

This approach was strongly criticised by the European Court of Human Rights (ECtHR). In unequivocal terms it declared that the standard of review before the domestic courts was:

> Placed so high that it effectively excluded any consideration by the domestic courts of the question of whether the interference with the applicants' rights answered a pressing social need or was proportionate to the national security and public order aims pursued…[56]

Irrationality simply did not afford enough protection to the rights of the applicants. In adopting irrationality as their standard of review the domestic courts had failed to consider the justifications for the policy. They had instead relied upon the unsubstantiated claim that allowing homosexuals to serve in the military jeopardised morale and operational effectiveness. They were 'confined'[57] to asking whether the policy was irrational and simply refused to address the issue of whether a fair balance had been struck between the rights of the applicants and competing public interests.

In contrast, in applying proportionality, the ECtHR placed great emphasis on the justifications behind the policy. It concluded that the government had not offered 'convincing and weighty'[58] reasons to justify the policy. There was no evidence to suggest that the presence of homosexuals threatened the effectiveness of the armed forces. Furthermore, the burden was on the government to provide this evidence. It was clear from the judgment in *Smith & Grady* that, in the context of Convention rights, irrationality was not a sufficiently robust standard of review. As Clayton argues, the ECtHR 'specifically rejected the notion that the *Wednesbury* doctrine adequately protects Convention rights.'[59] It was only proportionality, with its emphasis on justification, which provided a level of protection compatible with the ECHR.

Heeding the words of the ECtHR, the House of Lords in *Daly* emphasised the need for justification when the rights of an individual are limited. In language similar to that of the ECtHR in *Smith & Grady*, Lord Bingham stated the court must consider whether a limitation was as a 'necessary and proper response'[60] to the social need that is being addressed. In the context of the case he concluded that he could not accept the reasons put forward to justify the blanket ban on prisoners being present whilst their cells were searched.[61] In so doing, his Lordship adopted an approach that was far more rigorous

[55] ibid 558.
[56] *Smith & Grady* (n 51) para 138.
[57] ibid para 132.
[58] ibid para 105.
[59] Richard Clayton, 'Regaining a Sense of Proportion: the Human Rights Act and the Proportionality Principle' [2001] EHRLR 504, 509.
[60] *Daly* (n 44) [18].
[61] ibid [19].

than he and his colleagues had done in *Smith*. It was not sufficient that the policy fell within a range of reasonable responses open to the primary decision maker. The policy had to be justifiable, there had to be strong reasons supporting it. As such *Daly* must be regarded as a significant step away from *Wednesbury*, at least in the context of human rights. According to Hunt, the decision was a 'major landmark on the road to the development of a true 'culture of justification''.[62] Proportionality had been formally adopted as the standard of review in cases brought under the HRA. The requirement of justification that had been so obviously lacking under *Wednesbury* had been granted a prominent place within judicial review.

2.4 A New Constitutional Order?

The enactment of the HRA has taken us towards the culture of justification to which Mureinik aspired. By incorporating most of the ECHR into domestic law, the Act provides a bill of rights along the democratic model. Rights are not absolute principles over which the courts enjoy supreme interpretative authority. Rather the determination of rights may be regarded as a joint enterprise between the courts and the elected branches of government. The democratic organs of state are permitted a say in the protection of individual rights. It is this model that allows an examination of the justifiability of the decisions of the elected branches.

The courts have also adopted a new, more demanding approach to determining the acceptability of any rights limitation. In the context of human rights cases the courts have embraced the doctrine of proportionality. Unlike the highly restrictive *Wednesbury* irrationality, proportionality requires the judiciary to pay greater attention to the justifications behind any limitation of individual rights. Rather than asking whether the decision in question was one reasonably available to the primary decision maker, proportionality requires a more complex analysis of the rights at stake and any countervailing interests. The courts must determine whether any limitation is justifiable within the context of the ECHR, not whether it passes the high threshold of irrationality.

As such, the HRA has introduced a new constitutional order. The elected branches of government can no longer rely solely upon their democratic credentials or superior constitutional status when limiting Convention rights. Instead they are required to objectively justify their decisions, they must demonstrate that their decisions are justifiable not just rational. As a result of the HRA we have 'moved away from a model of majoritarian democracy'[63] towards a model with greater respect for the individual rights enshrined in the ECHR. The requirement of justifiability should ensure that such rights are not easily restricted.

[62] Hunt (n 48) 342.
[63] Jowell (n 49) 682.

3 A THREAT TO THE NEW ORDER: THE DISCRETIONARY AREA OF JUDGMENT

Having argued that the HRA has established a new constitutional order, in which primary decision makers are required to objectively justify any limitation of Convention rights, it is now time to examine how the judiciary have responded to this new order. This chapter is to criticise the concept of a 'discretionary area of judgment'[64] which features in the early case law under the HRA. Founded upon a rigid conception of the separation of powers and the same judicial culture of deference to the elected branches that spawned *Wednesbury*, the discretionary area of judgment for a primary decision maker has been used by the courts to avoid sufficiently robust review of rights infringement. In certain cases immediately after the enactment of the HRA the judiciary did not engage in the sophisticated analysis required by the doctrine of proportionality. Instead, the discretion of the primary decision maker was invoked by the courts to avoid an inquiry into the justifications of any limitation of Convention rights. As such, the concept of a discretionary area of judgment poses a significant threat to the culture of justification.

3.1 The Concept of the Discretionary Area of Judgment

When determining the validity of a restriction of a qualified Convention right the ECtHR has developed its own margin of appreciation doctrine. The court has recognised that:

> By reason of their direct and continuous contact with the vital forces of their countries, the national authorities are in principle better placed than an international court to evaluate local needs and conditions.[65]

As a supranational court, the ECtHR deems it necessary to accord the decisions of individual states, and their domestic institutions, a presumptive weight. As McBride explains, 'there is an assumption that those nearer the decisions might be better placed to assess the specific requirements of a situation.'[66] The ECtHR recognises that individual states are in a stronger position than itself to determine the appropriate response to domestic problems. The protection of Convention rights provided by the ECtHR is therefore 'subsidiary to the national systems safeguarding human rights.'[67] It is for states themselves to afford protection to Convention rights in the manner that best fits their own domestic situation.

The margin of appreciation doctrine has no place in domestic law. As Craig argues, the justification for the doctrine is 'integrally connected'[68] with the supranational nature of the ECtHR. It is founded upon a recognition that, as a supranational body, the court is

---

[64] Lester and Pannick, *Human Rights Law and Practice* (n 5) 73.
[65] *Buckley v UK* (1996) 23 EHRR 101, para 49.
[66] Jeremy McBride, 'Proportionality and the European Convention on Human Rights' in Evelyn Ellis (ed). *The Principle of Proportionality in the Laws of Europe* (Hart 1999) 29.
[67] *Handyside v UK* (1976) 1 EHRR 737, para 48.
[68] Paul Craig, 'The Courts, the Human Rights Act and Judicial Review' [2001] 117 LQR 589, 590.

limited in its ability to determine appropriate responses to domestic issues. The court is removed from 'local circumstance'[69] and therefore accords an element of deference to national institutions which are more likely to be in tune with the particular needs of that state. Such arguments are not applicable with respect to national courts. As Laws argues, the margin of appreciation will 'necessarily be inapt to the administration of the Convention in the domestic courts for the very reason that they are domestic.'[70] The doctrine is of supranational nature and is therefore inappropriate for adoption by domestic courts.

Despite the inappropriateness of the margin of appreciation doctrine in the context of domestic adjudication, it has been argued that national courts ought to develop a similar doctrine for judicial review under the HRA. Pannick contends that national courts should recognise a doctrine analogous to the margin of appreciation that accords a 'discretionary area of judgment in relation to policy decisions which the legislature, executive and public bodies are better placed than the judiciary to decide.'[71] He adds that the courts ought to recognise that in certain circumstances the elected branches of government are better placed to assess the needs of society and to carefully balance competing interests.[72] Just as the margin of appreciation doctrine is founded upon an assumption that national institutions are better placed to respond to particular domestic problems, the discretionary area of judgment assumes that Parliament and the executive occupy a superior position to the courts in the determination of questions of policy.

The danger with such an assumption, and consequently with the concept of a discretionary area of judgment, is that particular categories of Parliamentary and executive action can effectively be placed beyond the reach of judicial review. As Allan argues, the discretionary area of judgment is 'tantamount to a justiciability doctrine, premised on the idea that certain issues are not in any circumstances amenable to judicial determination.'[73] As the concept automatically assumes that the elected branches are better placed to make certain decisions, such as those relating to matters of policy, it encourages the judiciary to avoid any meaningful review of such decisions. As argued in the previous chapter, the HRA has replaced an assumption in favour of the superior status of the elected branches with a requirement that they justify any decision or action they take that limits Convention rights. That is the new constitutional order established by the HRA. The courts are charged with the responsibility of assessing such justifications. Failure to do so amounts to an abdication of judicial responsibility.

---

[69] ibid.
[70] John Laws, 'The Limitations of Human Rights' [1998] PL 254, 258.
[71] David Pannick, 'Principles of Interpretation of Convention Rights under the Human Rights Act and the Discretionary Area of Judgment' [1998] PL 545, 545.
[72] ibid 549.
[73] TRS Allan, 'Human Rights and Judicial Review: A Critique of 'Due Deference'' [2006] CLJ 671, 687.

The discretionary area of judgment serves as an alternative to thorough judicial scrutiny of the justifications offered in defence of limitations of qualified Convention rights. In the words of Edwards, adoption of the concept renders judicial review little more than a 'smell test'.[74] The mere presence of policy related matters causes the judiciary to avoid any meaningful analysis of the impugned decision. Where such issues are raised, the discretionary area of judgment acts as a convenient excuse allowing the courts to avoid issues of political controversy. The emphasis is placed upon the type of decision and not whether it is justified. As Hunt argues, the discretionary area of judgment regards certain areas of decision making as being 'within the realm of pure discretion'.[75] The courts are prevented from assessing the justifications behind decisions that fall within such areas of discretion. The discretionary area of judgment identifies certain spheres of decision making as being insulated from judicial review. Such an outcome is contrary to the culture of justification fostered by the HRA.

3.2 The Discretionary Area of Judgment in Action

The discussion thus far has been abstract. To demonstrate the effect of the discretionary area of judgment it is necessary to turn to the case law. Rather than being confined to the pages of academic journals, the discretionary area of judgment has been adopted by the judiciary in cases concerning issues of acute political controversy. An example of such an issue is immigration. This section is dedicated to an examination of some notable immigration cases where the discretionary area of judgment, if not explicitly named as such, has proved significant in the reasoning of the courts. The examination of the case law is intended to demonstrate that the concept has been used by the courts to avoid the more sophisticated judicial review required under the HRA. As argued above, the discretionary area of judgment negates the requirement that a primary decision maker objectively justifies the limitation of a Convention right, in contravention of the culture of justification.

An early manifestation of the discretionary area of judgment came in the case of *Rehman*[76] which combined the judicial hot potatoes of immigration and national security. The case centred on a deportation order made by the Home Secretary on the grounds that Rehman was involved in terrorist activity in India and therefore posed a threat to the UK's national security. Rehman challenged the order on the basis that he posed no such threat. He argued that his group, Markaz Dawa Al Irshad (MDI), had no intention to and could not target the UK. Consequently he and his group posed no threat to the security of the UK. Though not decided under the HRA, the case provides an example of how the concept of a discretionary area of judgment can act as a substitute for effective judicial scrutiny of an executive decision.

The key question on which the case hinged was whether Rehman did actually pose a threat to national security. If he did then the deportation order made under section

[74] Richard Edwards, 'Judicial Deference under the Human Rights Act' (2002) 65 MLR 859, 863.
[75] Hunt (n 48) 339.
[76] *Secretary of State for the Home Department v Rehman* [2001] UKHL 47, [2003] 1 AC 153.

3(5)(b) of the Immigration Act 1971 was lawful, if not, the deportation could not be justified. Despite the central importance of the question to the legality of the deportation, the House of Lords refused to confront it. Instead their Lordships granted the executive a wide discretionary area of judgment to determine whether Rehman was a threat to national security. Leading the way in this regard was Lord Hoffmann. He stated that:

> the question of whether something is in the interests of national security is not a question of law. It is a matter of judgment and policy. Under the constitution of the United Kingdom and most other countries, decisions as to whether something is or is not in the interests of national security are not a matter for judicial decision. They are entrusted to the executive.[77]

For Lord Hoffmann, the issue was not how the courts ought to approach questions of immigration and national security but whether such questions can "properly be decided by a judicial tribunal at all."[78] Whilst not entirely dismissing the court's ability to review determinations of national security, Lord Hoffmann's preferred course of action was to grant huge discretion to the executive in this field of decision making. In his postscript he argued that there was a need for the judiciary to "respect the decisions of ministers of the Crown" on questions of national security.[79] This was a category of decision making that warranted very little, if any, judicial interference. Lord Hoffmann and his colleagues refused to question the executive's assertion that Rehman's continued presence in the UK posed a threat to national security. Their Lordships sought no justification for the Home Secretary's claim.

Underlying Lord Hoffmann's reluctance to challenge the Home Secretary's decision to deport Rehman, was the discretionary area of judgment. For Lord Hoffmann, considerable discretion was to be accorded to the executive when making policy decisions. Elsewhere he argued that it was a "legal principle" that majority approval was necessary for the determination of policy questions.[80] This would appear to suggest that the courts, who are not democratically elected, are unable to consider any cases involving elements of policy. Such a position runs counter to the very purpose of the HRA. As Lord Steyn argues, in principle "there cannot be any no-go areas under the ECHR and for the rule of law."[81] Seeking to immunise certain categories of decision making, such as immigration, undermines the protection of individual rights that the HRA is supposed to facilitate. It does not require primary decision makers to objectively justify the limitation of Convention rights. It is the antithesis of the more stringent judicial review required by the culture of justification.

---

[77] ibid [50].
[78] ibid [52].
[79] ibid [62].
[80] *R (Prolife Alliance) v British Broadcasting Corporation* [2003] UKHL 23 [2004] 1 AC 185 [76].
[81] Lord Steyn, 'Deference: A Tangled Story' [2005] PL 346, 351.

The discretionary area of judgment was also pertinent in the judgment of Lord Phillips MR in *Mahmood*.[82] Mahmood challenged his removal from the UK on the basis that it constituted a disproportionate limitation of his right to family life under Article 8 of the Convention. In response, the Home Secretary argued that such a limitation was justified in the interests of maintaining firm immigration control. As the decision to remove Mahmood was made by the Home Secretary prior to the HRA coming into force, the issue arose as to whether the case should be decided under the Act. Disagreeing with his colleague Laws LJ, Lord Phillips MR believed that the case ought to be decided under the HRA as the Home Secretary had expressed the view that his decision was compatible with Article 8.[83] It was appropriate for the courts to assess this claim.

Unfortunately, Lord Phillips MR did not subject the Home Secretary's claim to much assessment. As in *Rehman*, the discretionary area of judgment proved influential. Lord Phillips MR relied upon the judgment of Lord Hope in *Kebilene*[84] where it was stated that:

> In some circumstances it will be appropriate for the courts to recognise that there is an area of judgment within which the judiciary will defer, on democratic grounds, to the considered opinion of the elected body…[85]

With this as his foundation Lord Phillips MR argued that, in cases such as *Mahmood*, there will often be "an area of discretion permitted to the executive"[86] before a decision can be said to breach the Convention. The practical consequence of adopting the discretionary area of judgment was that the standard of review implemented by Lord Phillips is simply inadequate in the context of the HRA. He stated that the courts had to ask "whether the decision-maker could reasonably have concluded that the interference was necessary"[87], granting an area of discretionary judgment in the process. This is not what is required under the Convention. The question is whether the limitation of the qualified right is objectively justified, whether it is proportionate.
With respect, the approach of Lord Phillips MR in *Mahmood* contradicts the culture of justification. It is couched in the language of *Wednesbury* unreasonableness which, as demonstrated in the previous chapter, has been declared inadequate for human rights review. The decision maker merely had to establish that is was reasonable to conclude that the limitation of a Convention right was necessary, not that it was actually necessary or proportionate. Rather than focussing upon the justifications behind the limitation of Mahmood's right to family life, Lord Phillips MR instead emphasised the discretion of the executive. He avoided any meaningful examination of the Home Secretary's decision. The level of judicial scrutiny was not sufficiently strong. As

[82] *R (Mahmood) v Secretary of State for the Home Department* [2001] 1 WLR 840 (CA).
[83] ibid [36].
[84] *R v Director of Public Prosecutions, ex p Kebilene* [2000] 2 AC 326 (HL).
[85] ibid 381.
[86] *Mahmood* (n 83) [38].
[87] ibid [40].

Edwards argues, "Scrutiny at this level is not of the rigour demanded by the HRA. It is, in fact, judicial avoidance."[88]

A similar approach was adopted by the Court of Appeal in *Farrakhan*.[89] Here the issue was the validity of the Home Secretary's decision to prevent the applicant from entering the UK. The Home Secretary argued that Farrakhan's exclusion from the country was justified as his presence was not conducive to the public good. In similarity to *Rehman*, the judiciary were confronted with a case concerning an assessment of a threat to public order and security. As Farrakhan's exclusion amounted to a limitation of his qualified right to freedom of expression, under Article 10 of the Convention, the court's emphasis ought to have been upon the justifications for the exclusion. Lord Phillips MR appeared to recognise this, agreeing that Farrakhan's challenge was a "reasons challenge."[90] The key question was whether the Home Secretary had sufficient objective reason to exclude Farrakhan from the UK. In other words, was the Home Secretary's claim that Farrakhan posed a threat to public order justified? The requirement of justification ought therefore to have been central to the court's decision making process.

It is useful to consider the approach of Turner J who found in favour of Farrakhan at first instance.[91] The Court of Appeal subsequently reversed his decision. His judgment placed a strong emphasis upon the need for the Home Secretary to objectively justify the decision to deny Farrakhan entry. Turner J stated that there had to be "substantial objective justification"[92] for the decision to limit Farrakhan's right to freedom of expression. Clearly he was of the opinion that the Home Secretary's contention that the applicant's presence in the UK would jeopardise public order had not been objectively justified. There was, in Turner J's view, a "complete absence of evidence"[93] that allowing Farrakhan to enter the country would pose any threat to public order. He concluded that the Home Secretary had simply not demonstrated that there was "more than a nominal risk that community relations would be likely to be endangered"[94] should Farrakhan be permitted to enter the country. As a result, Turner J found in favour of Farrakhan on the basis that there was no justification for denying him entry to the UK.

The judgment is significant for its refusal to adopt the discretionary area of judgment approach. Turner J, in contrast to the House of Lords in *Rehman*, did not regard immigration policy or determinations of risk as being beyond the scope of judicial review. On the contrary, the question of whether Farrakhan's presence in the UK did actually pose a threat to public order was of the utmost importance. He did not merely

[88] Edwards (n 75) 868.
[89] *R (Farrakhan) v Secretary of State for the Home Department* [2002] EWCA Civ 606, [2002] QB 1391.
[90] ibid [5].
[91] [2001] EWHC Admin 781.
[92] ibid [48].
[93] ibid [53].
[94] ibid.

accept the arguments advanced by the executive on the basis that this was a policy related matter. Instead he subjected such arguments to rigorous scrutiny to assess their veracity and determine whether they constituted the objective justification required for the limitation of a qualified Convention right. His approach was far more consistent with the new constitutional order established by the HRA than that adopted by the House of Lords in *Rehman*.

Unfortunately however Turner J's approach was not adopted by the Court of Appeal. The court endorsed the discretionary area of judgment and found in favour of the Home Secretary without examining the justifications behind Farrakhan's exclusion. In fact, Lord Phillips MR admitted that the Home Secretary "advanced no evidence to justify his decision".[95] Despite this declaration the court concluded that the "Secretary of State provided sufficient explanation for a decision that turned on his personal, informed, assessment of risk" and that the limitation of Farrakhan's right to freedom of expression was proportionate.[96] Lord Phillips MR and his colleagues were convinced of the proportionality of the exclusion despite acknowledging that there was no evidence to suggest Farrakhan's presence in the country threatened public order.

As in *Rehman*, the primary concern of the Court of Appeal in *Farrakhan* was the discretion that was to be accorded to the executive within the sphere of immigration policy. The court was preoccupied with ensuring that they did not interfere with the Home Secretary's decision stating that "the margin of appreciation or discretion accorded to the decision maker is all-important".[97] This was to ensure that the court "avoids substituting its own decision for that of the decision maker."[98] In its desperate attempt to avoid such a substitution the court failed to discharge its duty under the HRA, namely to ensure that any limitation of a Convention right is justified. Adoption of the margin of discretion acted as a replacement for judicial review of the decision to limit Farrakhan's right to freedom of expression. As Allan argues, "The protection of freedom of speech was notional: in practice, the minister's judgment was not subject to serious judicial scrutiny."[99] By ignoring the requirement of objective justification the court abandoned its own role in the new constitutional order established by the HRA. It did not require the limitation of Farrakhan's Article 10 right to be justified.

3.3 The End of the Discretionary Area of Judgment

As Ewing puts it, "The incorporation of Convention rights has to mean something more than simply new lyrics for old songs."[100] The concept of the discretionary area of judgment may be the new lyrics adopted by the judiciary in the immediate aftermath of incorporation. However, in practice it amounts to the same old songs of excessive judicial deference in cases of acute political controversy. Just like the *Wednesbury*

[95] *Farrakhan* EWCA (n 90) [60].
[96] ibid [79].
[97] ibid [67].
[98] ibid.
[99] Allan (n 74) 690.
[100] KD Ewing, 'The Futility of the Human Rights Act' [2004] PL 829, 852.

doctrine, the discretionary area of judgment is founded upon a rigid conception of the separation of powers that severely limits the role of the courts in the determination of rights. In desperately trying to avoid substituting its own judgment for that of the primary decision maker, the judiciary have on occasion avoided making any judgment whatsoever and have not required objective justification for the limitation of qualified Convention rights. As such they have failed to fulfil their obligations in the new constitutional order established by the HRA. The discretionary area of judgment ought to suffer the same fate as *Wednesbury* in the context of human rights. It must be consigned to history, an embodiment of the old constitutional order that existed prior to the HRA.

4 MAINTAINING THE NEW ORDER: JUSTIFIABILITY

The HRA has established a new constitutional order in which primary decision makers are obliged to justify any limitation of Convention rights. By adopting the concept of the discretionary area of judgment the judiciary failed to uphold this justificatory obligation. Having criticised the discretionary area of judgement, the question now arises as to how the courts ought to approach judicial review under the HRA. It must be asked how the courts are to provide adequate protection to the rights enshrined in the Convention whilst remaining sensitive to their own institutional limitations. In answering this question it is useful to return to the distinction made earlier between justified and justifiable.[101] When reviewing any limitation of Convention rights, the courts must consider whether such a limitation is *justifiable*, not whether it is justified. The court does not seek convergence between the impugned decision and its own view of how a particular problem ought to be resolved. Instead, the court's focus is upon the reasoning behind the limitation and whether it constitutes a plausible, but not necessarily the only, argument in defence of the limitation.

The first, and primary, duty of the courts is the application of proportionality. As discussed earlier, proportionality is the substantive standard of review that has been formally adopted in judicial review under the HRA. It allows for a critical examination of the reasons behind any limitation of Convention rights. However, in applying proportionality it is sometimes acceptable for the courts to show a degree of deference to the primary decision maker. Unlike the discretionary area of judgment, the extent of judicial deference is not predetermined by the subject matter of the impugned decision. Instead, the guiding principle is justifiability. The courts must ask whether deference is justifiable with reference to the particular context and factual matrix of the case at hand.

4.1 What is Deference?

Arguing that judges ought, on occasion, to show a degree of deference to the primary decision maker risks negating the requirement of objective justification. Advocating judicial deference to primary decision makers risks a repeat of the judicial inaction that was criticised in the previous chapter. This concern was raised by Lord Hoffmann in

[101] David Dyzenhaus, 'Law as Justification' (n 17) 27-28.

Prolife Alliance.[102] His Lordship expressed the view that deference had "overtones of servility, or perhaps gracious concession".[103] He believed deference to be an inappropriate description of the court's approach as it suggested an obsequious judiciary incapable of holding Parliament and the executive to account. A similar argument has been advanced by Trevor Allan for whom any doctrine of deference is "empty or pernicious."[104] According to Allan, a doctrine of deference is a "tool of judicial discretion"[105] that "threatens the coherence of constitutional rights adjudication".[106] The concern is that judicial deference weakens the protection of rights offered by the HRA. It may undermine the requirement that any limitation of Convention rights be justifiable.

Given what has been argued this is a pressing concern. If judicial deference undermines the justificatory burden placed upon primary decision makers then it will damage the new constitutional order established by the HRA. However, judicial deference does not have to be regarded in this light. The views expressed by Allan and Lord Hoffmann centre on a concern that deference may amount to the complete abdication of the judicial role in rights adjudication. However, if properly understood, deference does not encourage the courts to abandon any meaningful review of primary decisions as the discretionary area of judgment did. In the words of Kavanagh, deference is "partial"[107] and not "absolute or complete".[108] It does not prevent a court from embarking upon its own assessment of the issue at hand. Rather, deference is a more modest requirement that judges attribute weight to the primary decision where it is justifiable to do so.

Deference ought to be understood as an "institutional"[109] form of judicial restraint. Its emphasis is upon the "comparative merits and drawbacks of the judicial process as an institutional mechanism for solving problems."[110] Deference does not exclude the courts from problem solving but it does require judges to be attentive to their own limitations. As Young puts it, "there may be institutional reasons for accepting that the legislature or the executive is more likely to reach the correct answer."[111] The courts ought to accept that, in some circumstances, the elected branches are better placed to determine the necessity of a limitation of a Convention right. The courts still carry out a proportionality analysis. Proportionality is the substantive element of judicial review under the HRA. However, it is justifiable for the courts to accord weight to the determinations of a primary decision maker whose institutional features render them more likely than the courts to arrive at a reasoned outcome.

---

[102] *R (Prolife Alliance) v British Broadcasting Corporation* [2003] UKHL 23, [2004] 1 AC 185.
[103] ibid [75].
[104] Allan, 'Human Rights and Judicial Review' (n 74) 675.
[105] TRS Allan, 'Judicial Deference and Judicial Review: Legal Doctrine and Legal Theory' (2011) 127 LQR 96, 98.
[106] ibid 99.
[107] Aileen Kavanagh, *Constitutional Review under the UK Human Rights Act* (CUP 2009) 172.
[108] ibid.
[109] Jeff King, 'Institutional Approaches to Judicial Restraint' (2008) 28 OJLS 409.
[110] ibid 410.
[111] Alison Young, 'In Defence of Due Deference' (2009) 72 MLR 554, 562.

As Kavanagh argues, deference is a "rational response"[112] to uncertainty. Where a judge is unsure of a particular issue it is not an abdication of the judicial role to accord weight to the primary decision maker's own view. On the contrary, in circumstances of uncertainty, deference is a reasonable approach. An example of such a situation could be evaluating a potential threat to national security. The court is likely to be extremely uncertain of the extent of the threat posed by a particular individual or group. Determining the threat requires anticipation of future harms on the basis of credible intelligence, a difficult task for the courts who are generally ill equipped to make these judgments. Displaying a degree of deference to the executive's risk assessment is therefore justifiable. It does not amount to judicial inaction, as the discretionary area of judgment did. Rather it is an acknowledgement from the courts that they are uncertain of a particular issue and that other branches of government may be better equipped to make such judgments.

This was recognised by the majority of the House of Lords in the *Belmarsh Prison* case.[113] The UK government sought to derogate from the Convention to impose indefinite detention on foreign terror suspects. In determining the acceptability of such a derogation the House of Lords had to decide whether the UK faced a "public emergency threatening the life of the nation" as required by Article 15 of the Convention. On this issue, the court accorded a strong degree of deference to the executive. Lord Bingham argued that "great weight"[114] should be given to the Home Secretary's risk assessment. The decision required a "factual prediction"[115] of what people may do, making it "necessarily problematic."[116] His Lordship was uncertain as to the extent of the threat faced by the UK. It was therefore justifiable to accord strong weight to the executive's own assessments, especially given their superior access to intelligence.

Despite according strong weight to the executive's risk assessment, the court was still able to find in favour of the applicants. The case centred on the necessity of the detention, an issue on which the court correctly displayed far less deference. Deference is not applied in a blanket manner. The key question is whether deference is justifiable in a particular context. In some circumstances judicial deference is justifiable, in others it is not. It is therefore wrong to characterise deference as an abdication of the judicial role. As Kavanagh puts it, "deference is not anathema to the culture of justification".[117] It requires judges to accord an appropriate weight to primary decisions where such weight is justifiable, taking into account the particular context. Judges must be able to

[112] Aileen Kavanagh, 'Deference of Defiance?: The Limits of the Judicial Role in Constitutional Adjudication' in Grant Huscroft (ed), *Expounding the Constitution: Essays in Constitutional Theory* (CUP 2008) 208.
[113] *A and Others v Secretary of State for the Home Department* [2004] UKHL 56, [2005] 2 AC 68.
[114] ibid [29].
[115] ibid.
[116] ibid.
[117] Kavanagh, *Constitutional Review* (n 108) 174.

demonstrate that deference is appropriate. Deference cannot just be invoked as a means of avoiding controversial issues as the discretionary area of judgment was. The principle of justifiability serves as a restriction upon judicial deference, ensuring that the protection of Convention rights is not abandoned.

When deference is understood in this manner it is capable of withstanding the concern that it may amount to an abdication of the judicial role. It is not intended to exclude the courts from human rights adjudication. Rather it serves as a reminder that, as Gearty puts it:

> The judges make up an important branch of the state but it is no more than one branch, with there being two others, the legislative and executive, towards both of which the courts must show sensitivity and understanding.[118]

Judicial deference amounts to a recognition of the fact that the courts are not always best placed to determine the necessity of the limitation of a Convention right. The judiciary ought to recognise that in some circumstances the institutional characteristics of Parliament and the executive render their decisions worthy of weight. In such circumstances, judicial deference is justifiable.

4.2 Determining the Degree of Deference

Deference is the practice of attributing weight to a primary decision. The extent of the weight, and consequently the degree of deference, is to be determined by what is justifiable within the context of a particular case. What follows is a discussion of two such considerations that may lead a court to display deference. These are the relative expertise of the courts and the elected branches and the requirement of democratic legitimacy. It is to be argued that relative expertise is a justifiable reason for deference. However, democratic legitimacy is not. Deference on this ground undermines the new constitutional order established by the HRA.

4.3 Relative Expertise

There are situations in which the superior expertise of the elected branches renders a degree of judicial deference justifiable. As Lord Steyn argues, courts may owe deference to the elected branches when they are "institutionally better qualified to decide the matter".[119] The central concern here is the decision making capability of the courts in comparison to the primary decision maker. The emphasis is upon the "capacity"[120] of the courts to make the relevant decision. Issues of pertinence include the procedures of the courts and the access to important information that they enjoy. Acknowledging the limits of the court's decision making capacity does not negate the requirement that any limitation of a Convention right be objectively justifiable. Rather it is a rational recognition that another branch of government may be better placed to make a particular decision.

---

[118] Conor Gearty, *Principles of Human Rights Adjudication* (OUP 2004) 117.
[119] Steyn, (n 82) 350.
[120] Jeffrey Jowell, 'Judicial Deference and Human Rights: A Question of Competence' in Paul Craig and Richard Rawlings (eds), *Law and Administration in Europe* (OUP 2003) 73.

Kavanagh provides a useful example explaining the rationale behind deference on the ground of relative expertise. When consulting a doctor for advice regarding a medical condition we display deference to their opinion. This is not an unreasonable approach. On the contrary, such deference to the considered opinion of a medical practitioner is an entirely reasonable course of action. We recognise that they 'possess medical training and expertise we lack.'[121] In such a situation we are not precluded from forming our own opinion. We may have experienced the particular medical condition in the past and feel confident in managing it. However, regardless of our own experiences, we accord strong weight to the doctor's diagnosis and prescription. As Kavanagh puts it, the 'primary rationale'[122] for doing so is 'respect for the acknowledged superior claims or qualities of the other.'[123] We recognise that a doctor possesses greater medical expertise than ourselves. Consequently, a degree of deference to their opinion is entirely rational, even desirable.

In judicial review under the HRA, it is clear that on occasion the elected branches will possess greater expertise than the courts. In such circumstances it will be justifiable for judges to attach weight to the decisions of the elected branches, just as a patient would attach great weight to the medical opinion of their doctor. Unlike the discretionary area of judgment, this approach to deference does not automatically assume that a case centred on a particular subject warrants judicial restraint. Judges would not be permitted from abandoning any meaningful review of the primary decision simply on the basis of the type of decision being challenged. However, judges ought to recognise the limits of their decision making capability. As Keene argues, on some issues the court should 'attach considerable weight to the views of those who actually had expertise in such matters'.[124]

The key here is that judges ought to recognise their limitations with respect to *some issues*. As noted above, deference does not apply in a blanket manner, even when controversial political issues are at stake. Any decision making process under the HRA will consist of a number of steps and the degree of deference may vary for each. The degree of deference will depend upon what is justifiable on that particular part of the decision making process, depending upon the relative expertise of the courts and the primary decision maker. Consider again the situation in the *Belmarsh Prison* case where the executive sought to derogate from the Convention to deal with a perceived threat to national security.

[121] Kavanagh, 'Deference or Defiance?' (n 113) 187.
[122] ibid.
[123] ibid.
[124] Sir David Keene, 'Principles of Deference under the Human Rights Act' in Helen Fenwick, Gavin Phillipson and Roger Masterman (eds), *Judicial Reasoning under the UK Human Rights Act* (CUP 2007) 209.

In such a case the first issue to be determined is whether there is a threat to national security. On this first question, judicial deference to the executive is justifiable. As Jowell argues:

> there is no reason why the courts may not concede the superior intelligence gathering capacity of the executive to answer that question accurately – or at least more accurately than the courts.[125]

Here the courts ought to recognise the superior expertise of the executive in determining the extent of the threat. The executive has greater access to relevant intelligence. As such, the executive's assessment ought to carry great weight. That does not prevent the court from challenging such an assessment. Where the executive advances a risk assessment without any foundation the courts ought not to accept it. That was the approach of Turner J at first instance in *Farrakhan*.[126] He found against the Home Secretary on the basis that no evidence had been advanced supporting the claims that Farrakhan posed a threat to public order. However, executive assessments of the threat posed by a particular individual to security will generally attract a degree of deference on the ground of relative expertise. The executive is better placed to make such judgments. As such, according strong weight to those judgments is justifiable.

That is not the end of the story. Whilst executive risk assessments may attract a strong degree of judicial deference there are other considerations for the court which may not. Identifying a legitimate objective, such as protecting national security, is only one part of the proportionality inquiry. The other two components of proportionality, that the measures adopted are rationally connected to the objective and that the right is impaired no more than is necessary, are far less deserving of judicial deference. As Jowell argues, the necessity of the measure is the "crunch constitutional question"[127] in the new constitutional order. The primary decision maker does not enjoy any obvious superior expertise in the determination of these questions. Consequently, the same degree of deference is not justifiable.

That is why the House of Lords in *Belmarsh Prison* adopted a strong position on the requirement that executive detention be necessary to counter the threat to national security. Whilst according a strong degree of deference to the executive on the question of whether there was a threat to national security, their Lordships displayed no such deference in challenging the necessity of the measure. The executive may enjoy superior expertise in determining the extent of the threat posed to the country, but the courts must ensure that a right is infringed no more than is necessary. As Lord Nicholls stated, "The duty of the courts is to check that legislation and ministerial decisions do not overlook the human rights of persons adversely affected."[128] On this key issue, the

---

[125] Jeffrey Jowell, 'Judicial Deference: Servility, Civility or Institutional Capacity?' [2003] PL 592, 598.
[126] *Farrakhan* EWHC (n 90).
[127] Jowell, 'Servility, Civility or Institutional Capacity' (n 126) 598.
[128] *Belmarsh Prison* (n 114) [80].

necessity of any rights limitation, less deference is due to the elected branches. Determinations where the primary decision maker enjoys superior expertise warrant a degree of judicial deference. Others, where the primary decision maker enjoys no such advantage, are not deserving of judicial deference. Deference is not justifiable.

4.4 Democratic Legitimacy

As well as relative expertise, it has been argued that the superior democratic credentials of the elected branches are another factor that may warrant judicial deference. Hunt argues that a doctrine of deference ought "to have space for a proper role for democratic considerations, including a role for the democratic branches in the definition and furtherance of fundamental values."[129] When reviewing the decisions of the elected branches courts ought to be mindful of the fact that they are reviewing decisions made by democratically elected and publically accountable bodies. As the courts do not enjoy such characteristics they ought to act with restraint. Failing to do so would amount to a usurpation of the democratic process and would elevate the unelected judiciary to the role of primary decision maker.

It is difficult to see the use of allowing democratic legitimacy as a factor determining judicial deference. As King puts it "Once we have accepted it is relevant, the issue becomes what role it should play."[130] In comparison to the courts, the elected branches will always have greater democratic legitimacy. They are by nature elected and accountable to the public. Attributing weight to a primary decision on the basis of democratic legitimacy only appears to amount to a bland assertion that the courts ought to acknowledge their secondary role. Such a secondary role is already preserved by the distinction between justified and justifiable. Remember, the court's focus is upon the reasons behind the limitation of a Convention right and whether they constitute a plausible argument in defence of the limitation. The court does not seek convergence between the impugned decision and its own view of how a particular problem ought to be resolved. This maintains the secondary function of the judiciary. There is no need to introduce democratic legitimacy as a factor determining deference as respect for the primary role of the elected branches is already implicit within the judicial review process.

Allowing democratic legitimacy to determine the degree of deference adds little to the discussion. Doing so could also undermine the justificatory burden imposed upon primary decision makers when limiting Convention rights. As Jowell argues, if democratic legitimacy is itself a reason for judicial deference

> courts would be automatically required to defer, on constitutional grounds, on any occasion on which a qualified right was claimed to be sacrificed on the altar of public interest.[131]

[129] Hunt, (n 48) 350.
[130] King (n 110) 436.
[131] Jowell, 'A Question of Competence' (n 121) 73.

If democratic legitimacy warrants judicial deference, what is to stop the courts from always deferring to the elected branches? Kavanagh responds by arguing that as deference is a matter of degree, and not an "all-or-nothing matter", democratic legitimacy will never be determinative of the judicial decision.[132] Deference is indeed a matter of degree. It is a matter of attributing weight to the primary decision. But if democratic legitimacy is a basis for attributing such weight, the review of a primary decision will always be weighted in favour of the elected branches. This is what Kavanagh describes as "minimal deference"[133], the presumption in favour of the primary decision that ought always to be present in judicial review under the HRA.

Such an outcome is contrary to the new constitutional order established by the HRA. As has been argued, the HRA imposes a burden on primary decision makers to objectively justify any limitation of Convention rights. The language of the qualified rights is clear, any limitation must be "necessary in a democratic society." This unequivocally instructs primary decision makers to demonstrate the necessity, or in common parlance, the proportionality of any rights restriction. It is therefore unjustifiable to advocate a universal presumption in favour of the primary decision, as Kavanagh does. The elected branches must justify any limitation of Convention rights. Any attempt to water down that requirement by introducing presumptions in favour of primary decision makers undermines this justificatory burden and the protection afforded to Convention rights. Reducing the burden on the elected branches on the basis of their democratic credentials runs counter to the new constitutional order established by the HRA.

4.5 The Concept of Justifiability

Instead of the discretionary area of judgment, the courts ought to be guided by justifiability to ensure that they provide adequate protection to Convention rights whilst remaining sensitive to their own institutional limitations. Proportionality is the substantive basis for judicial review under the HRA and ought to be the primary concern of the courts. However, when applying proportionality it may be justifiable for the court to accord a degree of deference to the primary decision maker. Rather than being an abdication of the judicial role, deference is the modest practice of attaching weight to the primary decision. The court is not prevented from assessing the issue nor is it permitted to abandon its core responsibility of protecting Convention rights. Instead, deference requires the courts to recognise that the elected branches sometimes enjoy institutional advantages rendering them better placed to make certain determinations. In such circumstances it is justifiable for the courts to accord strong weight to the conclusions of the elected branches.

Whilst deference on the basis of relative expertise is justifiable, deference on the basis of democratic legitimacy is not. Firstly, there is no need for allowing deference on this

[132] Kavanagh, *Constitutional Review* (n 108) 191.
[133] ibid 181.

basis. The secondary nature of the judicial role is inherent within the judicial review process, particularly as courts examine whether any rights limitation is justifiable and not whether it is justified. The judiciary are not promoted to the position of primary decision maker. It is unnecessary to repeat this in the form of an official judicial doctrine. Secondly, allowing deference on the ground of democratic legitimacy introduces a systemic bias in favour of the elected branches. This is contrary to the new constitutional order established by the HRA which imposes a justificatory burden on primary decision makers when they limit Convention rights. That burden should not be weakened by creating a presumption in favour of the elected branches.

5 CONCLUSION

Returning to the question of how judges ought to review the actions and decisions of the elected branches under the HRA, the answer offered here centres upon *justifiability*. The new constitutional order established by the HRA imposes a justificatory burden upon the elected branches when limiting Convention rights. They must demonstrate that such a limitation is objectively justifiable, supported by rational and cogent argument. No longer can the elected branches rely upon their democratic character or superior constitutional status to justify the limitation of Convention rights. It is for the courts, through proportionality, to critically asses the justifications offered in defence of a rights limitation. The concept of the discretionary area of judgment negated this justificatory obligation. It allowed the courts to avoid meaningful review of primary decisions concerning issues of political controversy, such as immigration. By granting the elected branches considerable discretion in certain fields of decision making, the courts failed to discharge their duty of ensuring that any limitation of Convention right be justifiable.

The primary concern of the courts in judicial review under the HRA is proportionality. It is the substantive basis for review and allows for a probing examination of the justifications behind a limitation of Convention rights. However, in applying proportionality it may be justifiable for the courts to display a degree of deference to the primary decision maker. It will be justifiable where the primary decision maker enjoys an institutional advantage over the courts rendering them better placed to make a reasoned decision. Whilst deference on the basis of relative institutional expertise is justifiable, deference on the basis of democratic concerns is not. The secondary nature of the judicial role is maintained by the fact that the courts ask whether a limitation of a Convention right is justifiable, not whether it is justified. The courts consider whether the reasons behind a limitation constitute a plausible defence of that limitation, not whether they would have adopted the same course of action. Furthermore, attaching weight to the decisions of the elected branches on the basis of their democratic credentials introduces a systemic bias in their favour. This weakens the justificatory burden placed upon them when limiting Convention rights.

The question of how judges ought to review the actions and decisions of the elected branches under the HRA is a question full of difficulty. This essay has offered a view

as to how that question should be answered. It is a view that is, of course, subject to challenge. One such challenge could be that given the introduction of a justificatory burden upon primary decision makers there is no need for deference. The courts ought simply to concern themselves with proportionality and not risk weakening the justificatory burden by introducing considerations designed to restrict the judicial role. Trevor Allan may, albeit in a more sophisticated form, raise such an objection.

However, such an argument ignores the considerable disagreement over human rights, their scope and when it is acceptable to limit them. It is wrong to argue that the courts should simply apply the doctrine of proportionality as if this will always yield the correct answer. There are no correct answers, only those that are more justifiable than others. The courts have to be prepared to accommodate, and respect, different views as to the necessity of restricting a Convention right. That is particularly the case where those different views are held by bodies whose institutional advantages over the court render them better placed to make certain determinations. That said, the courts cannot abandon their role of critically assessing the justifications offered in defence of a rights limitation. It has been suggested that they ought to be guided by what is justifiable when displaying deference as a means of ensuring that the justificatory burden imposed upon primary decision makers is not abandoned altogether.

POWER SHARING IN THE EUROPEAN UNION: HAS COURT OF JUSTICE ACTIVISM CHANGED THE BALANCE?

*FIONA JAYNE CAMPBELL**

This article argues that the activism of the Court of Justice of the European Union has shifted the balance of power within the European Union (EU) in favour of the EU against the member states. It conducts this examination in light of three theoretical conceptions of the balance of power: the traditional model, the primacy model and the theory of constitutional pluralism. Firstly, the article examines the effect of the development and expansion by the court of the principle of supremacy on the balance of power. Secondly, it analyses how the principles of direct, indirect and incidental effect have been developed by the court to shift power towards the EU. Finally, it argues that the court has moved power to the EU through its decisions based on fundamental rights. It is submitted that, in light of this shift in power, the primacy model best describes the current power sharing arrangement within the EU as opposed to the traditional model or the theory of constitutional pluralism.

1 INTRODUCTION

The European Union (EU) is the product of a series of substantial developments since its birth in in the aftermath of World War II. Entered into by six western European countries, the European Coal and Steel Community, as it was called then, had a purely economic purpose; to integrate the coal and steel industries (which were at that time seen as strategic sectors of the economy)[1] in order to prevent another world war. Today, the EU's purpose is no longer purely economic and it is involved in a plethora of issues within the now 27 member states (MS), soon to be 28 with the accession of Croatia. As the EU is active in an increasing number of policy areas, it is worth considering the implications this has on its federal nature.

Whilst it is true that federalism lacks any precise definition,[2] this thesis will examine federalism using the definition provided by Neff and Fischer; that federalism relates to governmental power-sharing arrangements[3] between central and sub central government. Using this narrow formulation of federalism, this thesis will exclusively analyse whether, through activism from the Court of Justice (CJ), the power sharing arrangement between the EU and MS, and thus the federal nature of Europe, has changed.

Traditionally, the power sharing arrangement between the EU and MS has been dictated by the principle of conferral, whereby the CJ may only act within the powers the MS

*Fiona Jayne Campbell, Newcastle University, Law LLB Stage Three.

[1] Loukas Tsoukalis, *What Kind of Europe?* (Rev edn, Oxford University Press 2005) 13.

[2] Stephen C Neff and Thomas C Fisher, 'Some American Thoughts about European "Federalism"' (1995) 44 International and Comparative Law Quarterly 904, 905.

[3] ibid.

have conferred on it through the treaties. The principle of conferral is recognised as the most important protection of the MS statehood (and consequently their sovereignty).[4]

In line with the principle of conferral, the power framework of the EU takes the form of a hierarchical structure whereby power is allocated from the MS up to the EU. Accordingly, it is for the MS to decide which powers to transfer to the EU and not for the EU to decide which powers to take from the MS.[5] From now on, this form of power sharing arrangement will be referred to as the 'traditional model'.

In contrast, activist judgments from the CJ have often seen the Court, through expansive treaty interpretations, apply EU law in areas previously thought to be the exclusive concern of national law. In this way, it might be argued that CJ activism has steered the EU toward a different hierarchical power sharing agreement wherein power is determined by the EU, rather than the MS, in contrast to the traditional model. This power sharing arrangement, which accords priority to EU law over national law, is referred to by Dougan as the primacy model.[6]

Alternatively, however, Rene Barents contends that the power allocation within the EU takes a heterarchial structure which he labels constitutional pluralism. Under this model, the two legal orders overlap in certain areas and are competing in order to obtain the largest market share…the law of one order is twisting itself into the law of the other, while that latter is resisting.[7] Thus, evidence of MS successfully resisting CJ activism would support the existence of constitutional pluralism between the EU and MS.

This thesis will identify that the power structure within the EU has shifted from the traditional model and will explore whether this shift has resulted in a structure of constitutional pluralism between the EU and MS or whether the primacy model better reflects this relationship. The aim of this thesis is to discredit the argument in favour of constitutional pluralism and show that the power sharing arrangement between the EU and MS accords with the primacy model. Thus, it is contended that, through judicial activism, the CJ has engineered a movement away from the traditional model, according precedence to MS, and toward the primacy model, according precedence to the EU.

Given the definition of federalism provided by Neff and Fischer, this thesis examines part of the federalism debate since it explores which federalist power sharing agreement best describes the EU set up. However, it is important to stress that this thesis is not advocating that the CJ is driving the EU toward a United States of Europe. The intention

[4] Dieter Grimm, 'Defending Sovereign Statehood against Transforming the European Union into a State' (2009) 5 European Constitutional Law Review 353, 362.
[5] ibid 354.
[6] Michael Dougan, 'When Worlds Collide! Competing Visions of the Relationship Between Direct Effect and Supremacy' (2007) 44 Common Market Law Review 931, 946.
[7] Rene Barents, 'The Precedence of EU Law from the Perspective of Constitutional Pluralism' (2009) 5 European Constitutional Law Review 421, 442.

is to demonstrate that the power structure between the EU and the MS is changing and an arrangement has emerged which favours EU law over the law of MS.

This thesis will explore three of the principal ways that the CJ has achieved this power transfer from the national to the European level. In particular, the CJ has significantly shifted the power balance in Europe through the introduction of the doctrines of supremacy, direct effect and the recognition of fundamental rights (FR).

2 SUPREMACY

2.1 Introduction

This Chapter will argue that the CJ, through the introduction and development of the principle of supremacy, has ensured that EU law is supreme over national law. This furthers the argument that the primacy model most accurately reflects the current power sharing arrangement between the EU and MS.

Firstly, to demonstrate this, the legal foundations of supremacy must be considered since, if there is no legal basis for the doctrine in the Treaty, then the CJ can be taken to have acted outside of its conferred powers and thus the argument in favour of the primacy model is furthered.

Secondly, the resistance of MS against attempts by the CJ to broaden its jurisdiction will be considered. It will be identified that MS resistance is decreasing, enabling the CJ to maintain expansive treaty interpretations which, in practice, extend its power and challenge the existence of constitutional pluralism within the EU.

Finally, the case for constitutional pluralism will be considered. It will be identified that recent decisions of the CJ, particularly in relation to immigration, demonstrate the ability of the CJ to expand its powers within the EU, despite resistance. This calls the idea of constitutional pluralism into question and favours the primacy model.

2.2 CJ Activism in Developing the Supremacy Principle

The doctrine of supremacy was first introduced by the CJ in the case of *Flaminio Costa v ENEL.*[8] To accord with the principle of conferral, the CJ should have acted within the powers conferred on it by the MS through the treaty when introducing the supremacy principle. However, this does not seem to have been the case since the textual evidence provided by the Court in *Costa* was unpersuasive.

In an attempt to point to textual support from the treaties, the CJ turned to what is now Article 288 Treaty on the Functioning of the European Union (TFEU),[9] which states that regulations are directly applicable and are therefore integrated into national law

[8] Case C-6/64 *Flaminio Costa v ENEL* [1964] ECR I-585.

[9] Treaty of Lisbon amending the Treaty on European Union and the Treaty establishing the European Community, signed at Lisbon, 13 December 2007, Art 288 Treaty on the Functioning of the European Union (TFEU).

without the need for implementing measures. The CJ concluded that this Article would be meaningless if states could nullify the effect by subsequent inconsistent legislation.[10] However, reliance on this treaty provision may be criticised. Firstly, the scope of direct applicability is different from that of supremacy, limiting the extent to which the former can justify the latter. Indeed, whilst Article 288 TFEU[11] refers to directly applicable regulations, the CJ used the provision to establish a general principle of the supremacy of all binding EU law.[12] Secondly, supremacy operates independently to direct applicability since direct applicability itself does not resolve the priority between EU law and national law.[13]

Problematically, as has been identified by Hartley, there is no other source from which such a principle could have derived.[14] In this respect, it is unsurprising that the CJ has been accused of introducing 'judge made law violating the treaty'.[15]

It should be noted that the lack of textual support for supremacy in the Treaty establishing the European Economic Community (EEC Treaty)[16] was not an oversight. Despite Craig and De Burca's contention that the CJ's teleological arguments (discussed next) are stronger,[17] neither suggests that supremacy was, in the minds of the MS, a foregone conclusion before *Costa*.

First, the CJ stated 'the member states have limited their sovereign rights and created a body of law which binds their nationals and themselves'.[18] Thus, the CJ justified supremacy by arguing that the principle was both intended by the founding members of the EU when they drafted the treaties and understood by the new MS as a consequence of their membership. However, this reasoning is questionable since the CJ made no reference to the constitution of any particular MS to see whether such a transfer or limitation of sovereignty *was* contemplated by MS, or was possible in accordance with that constitution.[19] This criticism is persuasive, especially since it is evident that, in respect of the original MS, there was in fact no specific constitutional preparation for this CJ inspired development.[20]

[10] *Costa* (n 8) 593.
[11] Lisbon Treaty, Art 288 (n 9).
[12] Paul Craig and Grainne De Burca, *EU Law Text, Cases and Materials* (5th edn, Oxford University Press 2011) 258.
[13] ibid.
[14] Trevor C Hartley, 'The Constitutional Foundations of the European Union' (2001) 117 Law Quarterly Review 225, 242.
[15] Hans Heinrich Rupp, *'Grundgesetz und Europaischer Verfassungsvertrag'* (2005) Juristenzeitung 741, 744 as cited in Barents (n 8) 442.
[16] Treaty Establishing the European Economic Community, entered into force 1 January 1958
[17] Craig and De Burca (n 12) 258.
[18] *Costa* (n 8).
[19] Craig and De Burca (n 12) 258.
[20] Joseph Weiler, 'The Community System: the Dual Character of Supranationalism' (1981) 1 Yearbook of European Law 267, 276.

Secondly, the CJ argued that the very aims of the Treaty could not be achieved unless primacy was accorded to EU law.[21] The CJ stated 'the executive force of EU law cannot vary from one state to another in deference to subsequent domestic laws, without jeopardising the attainment of the objectives in the treaty set out in Article 5(2) and giving rise to discrimination, prohibited by Article 7.'[22] However, whilst it is a clear objective that EU law be applied uniformly across MS, uniformity is not sufficient justification for supremacy. This is demonstrated by the fact that the treaties contain a range of opt out clauses which allow national actors to deviate from EU law.[23] Moreover, Article 20 Treaty on European Union (TEU)[24] provides for enhanced cooperation, whereby EU law can develop between nine MS where there is not a sufficient voting threshold for general legislation.[25] Clearly, that MS are permitted by the treaty to integrate to different degrees implies that uniformity is not required in EU law.

Additionally, the CJ was not only activist in introducing the principle of supremacy, but has subsequently expanded the scope of the principle several times, to make it more broadly applicable. In the case of *Internationale Handelsgesellschaft*[26] the CJ ruled that EU law could not be challenged on the basis that it runs counter to FR as formulated by the constitution of that state, or to principles of a national constitutional structure.[27] Thus, the CJ explicitly claimed supremacy for EU law over even constitutional rules of national law. The CJ expanded the scope of supremacy even further in the case of *Simmenthal*,[28] holding that the supremacy of EU law applied to national law which pre dated and post-dated the EU law.[29] Indeed, the justification provided by the CJ in both instances mirrored the teleological arguments put forward in *Costa* and is therefore unpersuasive.

2.3 Supremacy as a Challenge to the Principle of Conferral

Given the importance that MS attach to the principle of conferral, it is somewhat surprising that the CJ in *Costa* decided in complete contradiction of the principle; assuming the supremacy of EU law without any treaty provision embodying their views.[30]

---

[21] Craig and De Burca (n 12) 258.
[22] *Costa* (n 8).
[23] Mattias Kumm, 'The Jurisprudence of Constitutional Conflict: Constitutional Supremacy in Europe Before and After the Constitutional Treaty' (2005) 11 European Law Journal 262, 303.
[24] Treaty of Lisbon amending the Treaty on European Union and the Treaty establishing the European Community, signed at Lisbon, 13 December 2007, Art 20 Treaty on European Union.
[25]Damian Chalmers, Gareth Davies and Giorgio Monti, *European Union Law Text and Materials* (2nd edn, Cambridge University Press 2010) 113.
[26]Case C-111/70 *Internationale Handelsgesellschaft v Einfuhr und Vorratselle fur Getreide und Futtermittel* [1970] ECR I-1125.
[27] ibid para 3.
[28] Case C-106/77 **Amministrazione delle Finanze dello Stato v Simmenthal SpA [1978] ECR I-629.**
[29] Craig and De Burca (n 12) 261.
[30] ibid 266.

One way in which some academics have mitigated the effects of *Costa* on MS sovereignty is demonstrated in Weiler's argument that the doctrine is bi-dimensional.[31] In line with this argument, the elaboration of the parameters of the doctrine by the CJ is only one aspect of the doctrine of supremacy. Rather, its success also depends on the second dimension: its incorporation into the constitutional order of the MS and its affirmation by their Supreme Court.[32]

What is most salient for the purposes of this thesis, is that there seems to be a difference between the power structure advocated through conferral and that of a bi-dimensional doctrine of supremacy. As identified in the introduction, the principle of conferral supports a hierarchical structure, whereby precedence is given to national law. By contrast, Weiler's theory accords well with Barents' contention that constitutional pluralism exists within the EU since a bi-dimensional doctrine of supremacy allows for a balance of power between the EU and MS, whereby MS may resist the imposition of supremacy.

However, as will be demonstrated throughout the remainder of this Chapter, the actual ability of MS to resist the invasion of EU law within their national legal orders has been limited by the MS themselves and where resistance has been attempted, it has failed. In light of this, the doctrine of supremacy is becoming increasingly one-dimensional, which instead supports the existence of the primacy model.

2.4 Supremacy as a Challenge to Constitutional Pluralism

2.4.1 A Limitation of MS Resistance to Supremacy, by MS

One challenge to the existence of constitutional pluralism in the EU is the increasing lack of resistance to supremacy shown by MS. In the wake of the introduction of the supremacy doctrine by the CJ, national constitutional courts asserted the right to review acts of the CJ when exercising the supremacy principle. This has been acknowledged as a signal by the MS that they have not ceded power.[33]

However, as will be discussed next, it would appear that the national constitutional courts are extremely reluctant to employ their reserved powers and have actually imposed limitations on the number of circumstances in which they are permitted to intervene. Indeed, such decreasing resistance by MS challenges the existence of constitutional pluralism in the EU and furthers the argument that the primacy model best reflects the power sharing agreement between the EU and MS.

This retreat in MS resistance can be demonstrated most effectively when considering the dilution of the German Constitutional Court's (GCC) resistance. Firstly, in the *Brunner*[34] case, the GCC developed the *ultra vires* review through which it claimed

[31] Weiler 'The Community System: the Dual Character of Supranationalism' (n 20), 275.
[32] ibid 276.
[33] Kumm (n 23) 281.
[34] *Brunner v The European Union Treaty* [1994] 1 CMLR 57.

jurisdiction to review acts of EU institutions to ensure they did not act outside of their conferred power when exercising the supremacy principle. Secondly, in *Internationale Handelsgesellschaft*[35] the GCC refused to recognise the supremacy of EU law which conflicted with basic rights enshrined in the German Constitution.[36] Finally, the GCC articulated that it would not accept supremacy where doing so would impinge on Germany's constitutional identity: the so called identity lock.[37]

Whilst Grimm argues that such decisions will increase the EU's mindfulness that its legitimacy depends largely on the willingness of MS and that it should be hesitant to exhaust this capital,[38] this seems unlikely since the GCC appears to have relaxed its position.

The relaxation of the *ultra vires* review by the GCC is primarily evident in the *Honeywell* ruling[39] where the Court determined that it would only intervene where there was a 'manifest' violation of competences and where the impugned act was 'highly significant' in the structure of competence between the MS and EU.[40] As the dissenting Judge Landau argued, this makes it excessively difficult to conclude that the EU has acted *ultra vires*.[41]

Similarly, the potency of the FR review has 'lost its bite'[42] following the 'bananas decision'[43] where the GCC restricted its competence to exercise FR review to situations where a general deterioration in the EU level of protection has been demonstrated,[44] thus excluding from its review individual cases of FR violations. Indeed, this decision erected such high hurdles that it has become improbable that the court will exercise this reserved control.[45]

Finally, whilst it remains to be seen how the identity lock is applied in subsequent case law,[46] in light of the retreat from the *ultra vires* and FR review, it is likely that the identity lock will be equally rarely enforced, if ever. Thus, the extent to which these forms of review really demonstrate that the MS have not ceded power is questionable. Although such review of CJ supremacy decisions may be asserted as a matter of theory, the practical situations in which such review will actually take place have been

[35] *Internationale Handelsgesellschaft v Einfuhr und Vorratselle fur Getreide und Futtermittel* [1974] 2 CMLR 540.
[36] Craig and De Burca (n 12) 275.
[37] ibid 279.
[38] Grimm (n 4) 372.
[39] 2 BvE, 2/08, 30 June 2009.
[40] ibid para 61as cited in Craig and De Burca (n 12) 279.
[41] Craig and De Burca (n 12) n30.
[42] Kumm (n 23) 294.
[43] 2 BvL, 1/97, 6 July 2000.
[44] Frank Hoffmeister, 'German Bundesverfassungsgericht: Alcan Decision of 17 February 2000; Constitutional Review of EC Regulation on Bananas, Decision of June 2000' (2001) 38 Common Market Law Review 791, 802.
[45] ibid.
[46] Craig and De Burca (n 12) 282.

significantly restricted by the MS themselves. Moreover, the argument that the ability of MS to review decisions of the CJ may influence the CJ in showing greater sensitivity to national constitutional concerns[47] is weak since in practice, there is little evidence to suggest such sensitivity on the part of the CJ.

In *Tobacco Advertising*,[48] for the first time, the CJ repealed an EU measure of political importance on the ground of lack of competence[49] after a challenge from the GCC. Indeed, this is undoubtedly demonstrative of courts in a relationship of constitutional pluralism, since it indicated that the GCC would exercise its jurisdiction in relevant circumstances, despite its previous perceived reluctance, and was a sign that the CJ might be beginning to take jurisdictional issues more seriously[50] as a result. Moreover, it was against this background that the doctrine of supremacy was said to be bi-dimensional,[51] as discussed previously in Section 2.3.

However, in the subsequent cases, the competence restricting elements of the *Tobacco Advertising* case have been contradicted or eroded away,[52] which suggests that there has been an increase in CJ activism and a decrease in MS resistance. What is most significant for the purposes of this thesis, is that this decrease in MS resistance challenges the existence of constitutional pluralism. Whilst Weiler contends that the principle of supremacy is bi-dimensional, and thus that constitutional pluralism exists within the EU, the increasing reluctance of MS to challenge CJ opinions suggest that the second dimension of the principle may be practically insignificant. This being the case, the argument in favour of the existence of the primacy model is furthered.

The next section will demonstrate that where resistance is attempted by MS, it has proven unsuccessful. That the MS have restricted their own ability to resist the invasion of EU law, in addition to the fact that any attempts of resistance have been largely unsuccessful, stands in contradiction to Barents' contention that constitutional pluralism exists in the EU. Rather, such lack of effective resistance supports the argument that the primacy model of power sharing exists between the EU and MS.

2.4.2 Unsuccessful MS Resistance

In support of the existence of constitutional pluralism, Barents identifies that recent developments in the case law of the CJ regarding national systems of direct taxation demonstrate that, due to strong national resistance, the invasion of EU law can be partially rolled back.[53] However, the extent to which this strongly supports Barents'

[47] ibid 300.
[48] Case C-376/98 *Germnay v Parliament and Council* [2000] ECR I-8419.
[49] Anneli Albi, 'Supremacy of EC Law in the New Member States: Bringing Parliaments into the Equation of "Co-operative Constitutionalism"' (2007) 3 European Constitutional Law Review 25, 30.
[50] Kumm (n 23) 296.
[51] Albi (n 49) 30.
[52] Dereck Wyatt, 'Community Competence to Regulate the Internal Market' in Michael Dougan and Samantha Currie (eds), *50 Years of the European Treaties : Looking Back and Thinking Forward* (Hart 2009) 93.
[53] Barents (n 7) 442.

argument is doubtful due to the implication that strong national resistance produced only a partial retreat by the CJ.

Furthermore, as demonstrated below, recent developments in the case law of the CJ regarding immigration law indicate how the principle of supremacy has enabled the CJ to declare EU immigration law (the free movement provisions), superior to domestic immigration law and thus extend the scope of EU competence, with failed MS resistance. This challenges the existence of constitutional pluralism and instead supports the existence of the primacy model between the EU and MS.

In the case of *Metock,*[54] the CJ required Ireland to accept that EU nationals with third country national family members, who were not already residing within the EU, would have to be permitted into Ireland on the basis of Directive 2004/38/EC.[55] This was previously assumed to only apply after the third country national family member had already been granted residency rights in an EU MS on the basis of national immigration rules rather than the EU's free movement rules (the prior lawful residence rule). Thus, third country national family members of EU residents would have to be permitted to exercise EU 'free movement' rights into the EU, thus bypassing national immigration laws entirely. This judgement has been criticised by O'Leary, who states that the CJ drove a coach and horse through the MS competence in the field of immigration.[56]

What is most salient for the purposes of this thesis is that it was the principle of supremacy which enabled the CJ to justify approaching the decision on the basis of EU law as opposed to national law, which was previously thought to enjoy exclusivity in this area.

After justifying that the EU had competence to regulate (as it did by Directive 2004/38/EC)[57] the entry and residence of third country nationals into the MS in which their family member had exercised free movement rights,[58] the CJ deduced that the arguments put forward by the Minister of Justice and by several governments, that MS retain exclusive competence in this area, must be rejected.[59] Thus, the CJ justified deciding *Metock* on grounds of EU law, due to the existence of Directive 2004/38/EC which, it stated, precluded legislation of a MS which is based on the prior lawful residence rule.[60] This meant that the Directive in question applied in preference to

[54] Case C-127/08 *Blaise Beheten Metock and others v Minister for Justice, Equality and Law Reform* [2008] ECR I-6241.
[55] Directive 2004/38/EC of the European parliament and of the council of 29 April 2004 on the right of citizens of the Union and their family members to move and reside freely within the treaty of the Member States amending Regulation (EEC) No 1612/68 and repealing Directives 64/221/EEC, 68/390/EEC, 72/194/EEC, 73/148/EEC, 75/34/EEC, 75/35/EEC, 90/364/EEC, 90/365/EEC and 93/96/EEC OJ L158.
[56] Siofra O'Leary, 'The Past, Present and Future of the Purely Internal Rule in EU Law' (2009) 44 Irish Jurist 13,34.
[57] Directive 2003/38/EC (n 55).
[58] *Metock* (n 54) Para 65.
[59] ibid Para 66.
[60] ibid Para 80.

conflicting national law which clearly demonstrates that it is the doctrine of supremacy which underlies the CJ judgment in *Metock*. Despite this, the CJ in *Metock* never explicitly referred to the supremacy principle. This is significant since it suggests that the principle is so well accepted that, even when it is used controversially to expand competence, there is no need for its direct mention.

Moreover, that several governments put forward arguments during *Metock* in order to resist competence of EU law over national border control, illustrates the political salience of the case.[61] It is striking for the purposes of this thesis that the CJ gave 'short shrift'[62] to the arguments of the MS and nevertheless applied EU law, leaving MS powerless over the matter.[63] The failure of such resistance by MS and the lack of any effective resistance after the judgment, significantly questions Barents' contention that constitutional pluralism exists between the EU and MS. Instead, the ability of the CJ to use supremacy to deliver such controversial judgments favours the existence of the primacy model.

2.5 Declaration 17: Challenging Constitutional Pluralism

Barents' justification for advocating constitutional pluralism is that there is a divergence of views between the CJ and national courts as to the power structure in Europe.[64] In line with this, a merging of views could provide a sound basis to discredit the existence of constitutional pluralism. Indeed, such a merging of views could be seen to have taken place in light of the Lisbon Treaty.[65]

Whilst the rejected Constitutional Treaty included provisions on supremacy from Article 1-6, when the Treaty was re-drafted to form the Lisbon Treaty,[66] these articles were cut out and replaced with a declaration.

On the one hand, it might be argued that dropping the primacy clause from the Lisbon Treaty[67] was unwise because its removal might cause some national courts to doubt the continuing validity of the principle.[68] However, as Craig and De Burca have stated, this is unlikely.[69] In fact, Declaration 17 on supremacy can be seen as revolutionary,[70] since it is the first time that the *Costa* case law has been explicitly endorsed and ratified by all MS.[71] Moreover, the Declaration states that primacy takes effect 'in accordance with

---

[61] Cathryn Costello, '*Metock*: Free Movement and "Normal Family Life" in the Union' (2009) 46 Common Market Law Review 587, 597.
[62] Paul Craig, 'Once Upon a Time in the West: Direct Effect and the Federalisation of EEC Law' (1992) 12 Oxford Journal of Legal Studies 453, 468.
[63] Samantha Currie, 'Accelerated Justice or a Step Too Far? Residence Rights of Non-EU Family Members and the Court's Ruling in Metock' (2009) 34 European Law Review 310.
[64] Barents (n 7) 444.
[65] Treaty of Lisbon amending the Treaty on European Union and the Treaty establishing the European Community, signed at Lisbon, 13 December 2007.
[66] ibid.
[67] ibid.
[68] Craig and De Burca (n 12) 266.
[69] ibid.
[70] Declaration Concerning Primacy, attached to the Lisbon Treaty.
[71] Chalmers, Davies and Monti (n 25) 188.

well settled case law of the CJ…under the conditions laid down by the said case law'.[72] This is likely to give rise to a construction that EU law has primacy over all national law including national constitutional law[73] since that reflects the position of the CJ.[74]

Thus, in light of Declaration 17, the primacy of EU law can no longer be relegated as being merely the view of the CJ since MS have expressly accepted the CJ's view that EU law is supreme.[75] Thus, the foundations of Barents' argument may be called into question and the argument that the primacy model exists between the EU and MS is strengthened.

2.6 Chapter Summary

To conclude, it is clear that, when asserting the doctrine of supremacy in *Costa*, the CJ acted outside of its conferred powers and contrary to the intentions of the MS.
Whilst this challenged the traditional model, some argued that the supremacy doctrine was bi-dimensional and, therefore, national courts could resist the invasion of EU law into the national law that supports the existence of constitutional pluralism.

However, as has been identified, the MS have limited their ability to resist the supremacy of EU law. Moreover, it has been demonstrated that, where MS have attempted to resist the advances of the CJ claiming supremacy in new areas of law, and thus extending EU competence, these have been unsuccessful. Moreover, *Metock* demonstrated that supremacy has paved the way for controversial judgments which accord priority to EU law over national law. This undoubtedly supports the existence of the primacy model.

In light of this, the doctrine is now, as a matter of practice, one dimensional as the CJ increasingly claims EU law supreme in different areas. Thus, Barents' theory is open to challenge, especially in light of the apparent merging of views between the CJ and MS. Thus, the argument that the CJ has driven the EU toward the primacy model is furthered.

3 THE ENFORCEMENT OF EU LAW: DIRECT, INDIRECT AND INCIDENTAL EFFECT

3.1 Introduction

It has been recognised that direct effect, whereby European individuals may enforce EU law in national courts, is one of the most significant achievements of the CJ.[76] Moreover, as one of the principal tools through which EU law produces independent effects within national orders,[77] it is evident that, like supremacy, the doctrine of direct

[72] Declaration Concerning Primacy (n 70).
[73] Craig and De Burca (n 12) 266.
[74] *Internationalle Handelsgesellschaft v Einfuhr und Vorratselle fur Getreide und Futtermittel* (n 22).
[75] Chalmers, Davies and Monti (n 25) 188.
[76] Craig 'Once Upon a Time in the West' (n 62) 453.
[77] Dougan (n 6) 932.

effect was designed to ensure the hierarchical supremacy of EU law over national law.[78] The net result of the two doctrines means that, not only must national courts apply EU law through direct effect, but EU law supersedes any conflicting national law due to the supremacy principle. Therefore, EU law emerges, in practice, as the law of the land.[79] This clearly by-passes state sovereignty and involves a concentration of power in the hands of the CJ.

Over time, the doctrine of direct effect has developed and expanded to cover an increasing number of sources of EU law: from regulations to treaty provisions and directives. Thus, as the CJ claims hierarchical superiority for EU law over national law in so many different areas, the argument that the CJ has advanced the relationship between the EU and MS toward the primacy model is strengthened.

This Chapter will consider the extent to which the emergence and development of direct effect supports the claim that the CJ, through activist judgments, has engineered the development of the primacy model between the EU and MS. First, the introduction of direct effect and its application to treaty provisions, regulations and decisions will be considered. Finally, and most controversially, the application of the doctrine to directives will be discussed.

3.2 CJ Activism in Developing Direct Effect

Direct effect was first introduced in the case of *Van Gend en Loos* wherein a preliminary reference was sent to the CJ from the Belgian Government inquiring whether or not a treaty provision had direct application before national courts.[80] Against this backdrop, observations were submitted by the Belgian, German and Netherlands Governments which provide a useful indication of how the signatories to the EEC Treaty[81] perceived the obligations they had undertaken therein.[82]

For example, the Dutch Government argued that direct effect would contradict the intention of those creating the Treaty[83] and all three governments argued that the forms of action created under the then Article 169 and 170 EEC Treaty[84] were sufficient.[85] Such strong intervention made by half of the existing MS at the time, is surely indicative that the idea of direct effect did not accord with the obligations they assumed when creating the Treaty.[86] Moreover, this resistance to the development of direct effect is characteristic of Barents' model of constitutional pluralism.

---

[78] Craig 'Once Upon a Time in the West' (n 62) 466.
[79]Sacha Prechal, 'Does Direct Effect Still Matter?' (2000) 37 Common Market Law Review 1047, 1047.
[80] Case C-26/62 *NV Algemene Transporten Expeditie Onderneming van Gend en Loos v Nederlandse Administratie der Balastingen* [1963] ECR I-1.
[81] (n 16).
[82] Craig, 'Once Upon a Time in the West', (n 62) 458.
[83] Craig and De Burca (n 12) 184.
[84] (n 16).
[85] Craig, 'Once Upon a Time in the West' (n 62) 458.
[86] Craig and De Burca (n 12) 185.

Nevertheless, these arguments advanced by the MS were rejected by the CJ, which held that treaty articles could have vertical direct effect and could therefore be relied upon by individuals in a case against the state. Indeed, that MS resistance was unsuccessful challenges the existence of constitutional pluralism and therefore advances the argument in favour of the existence of the primacy model.

To justify its decision, the CJ put forward a teleological argument which was designed to directly challenge the Dutch Government's contention that the Treaty was the same as any international treaty and therefore could not have direct effect.[87] Thus, the CJs claim that the EU had created 'a new legal order of international law for which MS have limited their sovereign rights'[88] was a vital step in advancing direct effect.[89]

Secondly, the CJ referred to the preliminary ruling procedure set out in the treaty under what is now Article 267 TFEU.[90] The CJ argued that through this procedure, states had acknowledged that EU law could be invoked in national courts.[91] Indeed, this textual evidence for direct effect is not particularly strong.[92]

In terms of the scope of direct effect, the CJ in *Van Gend en Loos* established initial conditions for a treaty article to have direct which have been developed over the years. It is now settled that provided a provision is clear, precise and unconditional, it will be accorded direct effect.[93] Moreover, it is clear that treaty articles can also have horizontal direct effect so as to impose obligations on private parties vis-à-vis one-another.[94] Indeed, this substantially increases the CJ's market share of competence which advances the arguments in favour of the existence of the primacy model.

Additionally, the reach of both vertical and horizontal direct effect has been extended to regulations, in the *Slaughtered Cow*[95] case, and Decisions, in *Grad*,[96] which also increases CJ's market share of competence. However, as will be discussed, the situation with regard to Directives is more complicated.

3.3 The Vertical Direct Effect of Directives

In accordance with Article 288 TFEU,[97] MS have a duty to implement directives into their national legal system and the end result in each MS must be the same, although MS are granted discretion as to the further implementing measures required.[98] Once

[87] ibid 184.
[88] *Van Gend en Loos* (n 80).
[89] Craig, 'Once Upon a Time in the West' (n 62) 459.
[90] Lisbon Treaty (n 8), Art 267 Treaty on the Functioning of European Union.
[91] Craig, 'Once Upon a Time in the West' (n 62) 460.
[92] Craig and De Burca (n 12) 185.
[93] Orcun Senyucel, 'The Direct Effect of Community Directives: the Effect of the Unilever Judgement' (2005) Ankara Law Review 81, 83.
[94] Craig and De Burca (n 12) 189.
[95] Case C-39/72 *Commission of the European Communities v Italian Republic* [1973] ECR I-101.
[96] Case C-9/70 *Franz Grad v Finanzamt Traunstein* [1970] ECR I-825.
[97] Lisbon Treaty, Art 288 TFEU (n 8).
[98] Craig and De Burca (n 12) 192.

implemented through the national law, individuals in national courts may rely upon directives. The issue of direct effect arises where MS do not implement, or implement incorrectly, the directive.

Since directives allow for such discretion in implementation, it is difficult to reconcile their nature with the criteria established for direct effect[99] to apply; that they should be clear, precise and unconditional. Nevertheless, the vertical direct effect of directives was announced by the CJ in *Van Duyn.*[100] Indeed, the effectiveness objective is uppermost in the reasoning provided by the CJ[101] as the Court stated that to deny direct effect would ultimately result in the weakening of the useful effect of directives.[102]

The second justification was provided in response to the UK Government's argument that Article 288 TFEU[103] distinguishes between regulations and directives since, unlike regulations, directives are not directly applicable and therefore should not be accorded direct effect.[104] Indeed, the concept of direct effect of directives is not implicit in the wording of Article 288 TFEU.[105][106] Nevertheless, the CJ decided that, even though directives lack direct applicability, this did not preclude them from having similar effects to regulations.[107]

Moreover, the CJ repeated the argument advanced in *Van Gend en Loos*; that the preliminary reference procedure under Article 267 TFEU[108] allows national courts to refer questions to the CJ regarding directives, which implies that such acts can be invoked by individuals before national courts. As noted above in Section 3.2, this reasoning is weak. In fact, the most convincing rationale for the vertical direct effect of directives was developed in the *Ratti*[109] case and is based on an estoppel argument. Here, an unimplemented directive is binding against the MS since they should have implemented the directive.

The vertical direct effect of directives was met with substantial resistance by national courts.[110] In *Ministere de L'Interieur,*[111] the French Conseil D'Etat concluded that since national authorities remain exclusively competent to decide on the method of

[99] Craig and de Burca (n 12) 192.
[100] Case C-41/74 *Van Duyn v Home Office* [1974] ECR I-1337.
[101] Alan Dashwood, 'From Van Duyn to Mangold via Marshall: Reducing Direct Effect to Absurdity' (2006-2007) 9 Cambridge Yearbook of European Legal Studies 81, 85.
[102] *Van Duyn* (n 100) para 12.
[103] (n 9).
[104] Craig 'Once Upon a Time in the West' (n 62) 474.
[105] Deidre Curtin, 'The Province of Government: Delimiting the Direct Effect of Directives in the Common Law Context' (1990) 15 European Law Review 195, 196.
[106] (n 9).
[107] *Van Duyn* (n 100) para 12.
[108] (n 90).
[109] Case C-148/78 *Pubblico Ministero v Tullio Ratti* [1979] ECR I-1629.
[110] Craig, 'Once Upon a Time in the West' (n 62) 476.
[111] December 22 1978, *Ministere de l'Interieur c. Cohn-Bendit (1979) RTDE 157.*

implementation, directives may not be invoked by an individual in a national court.[112] Such opposition by national courts is significant since it may have been a substantial contributing factor to the CJ's denial of horizontal direct effect and may therefore lend force to Barents' theory of constitutional pluralism, as will be addressed later in Section 3.4.

3.4 The Marshall Rule: No Horizontal Effect of Directives

Following the trend of judicial activism of the CJ between 1960 and 1980,[113] it might have been expected that the natural next step for the Court was to grant horizontal direct effect to directives. However, in what has been described by Curtin as a turn to judicial minimalism,[114] the CJ in *Marshall*[115] articulated that directives are incapable of producing horizontal direct effect.

The judgment in *Marshall* was based on a narrow literal interpretation of Article 288 TFEU[116][117] since it considered that the phrase 'each MS to which it is addressed'[118] meant that directives could only create obligations for the state and not, therefore, for individuals. In *Vaneetveld*,[119] AG Jacobs criticised this reasoning in *Marshall*[120] for its formal nature which does not sit well with the purposive approach adopted in other areas.[121] Indeed, Craig has argued that we should cease to pretend that the answer is determined by textual argument since it is equally unlikely that the treaty framers imagined individuals to derive rights from the treaty or from regulations,[122] which are both accorded horizontal direct effect. Moreover, this textual reasoning is not actually convincing in black letter terms since it is more likely that the reference to MS was intended to address which MS were bound and not intended to touch on the issue of who within a state can be bound.[123]

Finally, whilst the estoppel argument articulated in *Ratti* is another basis on which to argue that directives should not be accorded horizontal direct effect,[124] as will be seen

---

[112] Pierre Pescatore, 'The Doctrine of "Direct Effect": An Infant Disease of Community Law' (1983) 8 European Law Review 155, 169.
[113] Curtin (n 105) 195.
[114] ibid.
[115] Case C-152/84 **M. H. Marshall v Southampton and South-West Hampshire Area Health Authority (Teaching)** ***[1968] ECR I-723.***
[116] Curtin (n 105) 196.
[117] (n 9).
[118] ibid.
[119] Case C-316/93 *Nicole Vaneetveld v Le Foyer SA and Le Foyer SA v Fédération des Mutualités Socialistes et Syndicales de la Province de Liège* [1994] ECR *I-00763.*
[120] Takis Tridimas, 'Horizontal Effect of Directives: a Missed Opportunity' (1994) 19 European Law Review 621, 627.
[121] Paul Craig, 'Directives: Direct Effect, Indirect Effect and the Construction of National Legislation' (1997) 22 European Law Review 519, 520.
[122] Paul Craig, 'The Legal Effect of Directives: Policy, Rules and Exceptions' (2009) 34 European Law Review 349, 352.
[123] ibid 352.
[124] Nick Maltby, 'Marleasing: What is all the Fuss About?' (1993) 109 Law Quarterly Review 301, 310.

in Section 3.5.1, the strength of this argument is somewhat diminished through the ECJ's broad interpretation of the state.

Indeed, as recognised in Section 3.3, in the absence of any convincing reason as to why the horizontal direct effect of directives should be denied, the reluctance of national courts to accept the vertical direct effect of directives was no doubt the cause of the CJ's decision in *Marshall*.[125] In this context, it might be said that the CJ has accepted limitations to the scope of its power to enforce EU law in the legal orders of the MS.

This acceptance of limitations undoubtedly supports the existence of constitutional pluralism within the EU. Indeed, as recognised by Tridimas, the ruling in *Marshall* compromises the primacy of EU law since it gives precedent to the very interests which the notion of direct effect was developed to challenge; those of the recalcitrant MS.[126] Nevertheless, the CJ subsequently sought to limit the ruling in *Marshall* in three categories of case law which will now be discussed in turn.

3.5 Circumventing the Marshall Rule

3.5.1 Broad Notion of the State

Since, in accordance with *Marshall*, directives may only impose obligations on the state, and not individuals, a definition of what constitutes the state is necessary. In *Foster*,[127] the CJ developed a broad interpretation of the state[128] and concluded that the state included bodies that are subject to the authority or control of the state or have special powers pursuant to their relationship with the state.[129] Consequently, the estoppel argument is deprived of much of its explanatory power,[130] since to follow the literal interpretation of Article 288 TFEU[131] would result in a narrow conception of the state.[132] Moreover, since the *Foster* ruling captures bodies with no greater power over the implementation of directives than private individuals,[133] such situations are clearly akin to horizontal direct effect. Thus, the CJ has circumvented the *Marshall* ruling and increased the scope of EU law in horizontal cases.

3.5.2 Indirect Effect

To further curb the effects of *Marshall*, the CJ imposed a requirement that national courts interpret national law in light of unimplemented directives. This is known as indirect effect and is classed as the most important qualification[134] to the *Marshall* ruling.

---

[125] Trevor C Hartley, *The Foundations of European Union Law* (Oxford University Press 2010) 234.
[126] Tridimas ''Horizontal Effect of Directives: a Missed Opportunity' (n 120) 633.
[127] Case C-188/89 *Foster and Others v British Gas plc* [1990] ECR I-3313.
[128] Miriam Lenz, Dora Sif Tynes and Lorna Young, 'Horizontal What? Back to Basics' (2000) 25 European Law Review 509, 513.
[129] *Foster* (n 127) para 20.
[130] Dashwood (n 101) 87.
[131] Lisbon Treaty, Art 288 TFEU (n 9).
[132] Craig, 'The Legal Effect of Directives: Policy, Rules and Exceptions' (n 122) 356.
[133] ibid.
[134] ibid 357.

The interpretative obligation, first articulated in *Van Duyn,* was expanded in *Marleasing*[135] which confirmed that it applies to national law pre and post dating the directive.[136] The effect of the CJs ruling in *Marleasing* was to enable the defendant company to successfully invoke the directive in its defence and defeat the claimant's action which was based on pre-existing national law.[137] Thus, as has been identified by Craig, this has the same results as if the directive had been accorded horizontal direct effect.[138]

Some maintain that this process of interpretation is demonstrative of a kind of heterarchial structure,[139] and thus constitutional pluralism, in the EU. Amstutz explains this position by articulating that the interpretative obligation is not a command to take over foreign law, but merely brings about a duty for the MS to rearrange their civil norms to the EU legal order.[140] Thus, where national law is interpreted in conformity with a directive, the individual is relying on the national law and not the directive in the proceedings. This is in contrast to direct effect, whereby the individual relies on the directive itself.

However, in practice this distinction makes little difference.[141] Indeed, whilst in conceptual terms, the interpretation is given effect through national law, it is nonetheless EU law that orchestrates the entire inquiry since it is the CJ that ultimately decides the types of cases to which the obligation applies.[142] Moreover, in reading the CJs interpretation of a directive into non-implementing national legislation, the national courts are transferring power to the CJ to dictate the domestic effect of EU laws that have no direct effect[143] which supports the existence of the primacy model.

Nevertheless, Amstutz furthers his argument by pointing to the fact that in *Marleasing*, the interpretative obligation was said to apply 'so far as is possible'[144] which he submits reduces the rigour of the requirement[145] since the national courts may act as regulators in the process of incorporating EU private law positions into the national legal

[135] Case C-106/89 *Marleasing SA v La Comercial Internacionale de Alimentacion SA*[1990] ECR *I-04135.*
[136] ibid para 8.
[137] Sara Drake, 'Twenty Years After Von Colson: the Impact of 'Indirect Effect' on the Protection of the Individual's Community rights' (2005) 30 European Law Review 329, 336.
[138] Craig, 'Once Upon a Time in the West' (n 62) 468.
[139] Marc Amstutz, 'In-Between Worlds: Marleasing and the Emergence of Interlegality in Legal Reasoning' (2005) 11 European Law Journal 766, 768.
[140]ibid 769.
[141] Takis Tridimas, 'Black, White and Shades of Grey: Horizontality of Directives Revisited' (2002) 21 Yearbook of European Law (2001-2002) 327, 348.
[142] Craig, 'The Legal Effect of Directives: Policy, Rules and Exceptions' (n 122) 369.
[143] Grainne De Burca, 'Giving Effect to European Community Directives' (1992) 55 Modern Law Review 215, 224.
[144] *Marleasing* (n 135) para 8.
[145] Amstutz (n 139) 771.

discourse.[146] Dougan has also argued that the *contra legem* exception stands in flat contradiction to the primacy model[147] whereby EU law is hierarchically supreme.

However, it is evident that the national courts are put under very strong pressure to reach an interpretation of existing national law consistent with the directive in question.[148] This is especially so in light of *Pfeiffer*[149] which gave the interpretative obligation a new level of urgency.[150] Indeed, the judgment repeated that the national court must do 'whatever lies within its jurisdiction'[151] to achieve an interpretation consistent with the directive. Consequently, it seems unlikely that the qualification 'as far as possible' in *Marleasing* will have much practical application since, in reality, little scope is left for an interpretative role by the national courts.[152]

In light of this, the approach of those favouring constitutional pluralism can be challenged and the argument in favour of the existence of the primacy model is furthered.

3.5.3 Incidental Effect

Another area of case law in which the CJ appears to have mitigated the effect of *Marshall*, has been termed the incidental horizontal effect of directives.[153] In *CIA*[154] the CJ allowed an unimplemented directive to be relied upon in proceedings between private parties before their national courts.[155] The case concerned Directive 83/189[156] which exclusively concerns the relationship between the Commission and MS and imposes a duty on the MS to notify the Commission on technical standards before they are adopted so that the Commission can verify their compatibility with the free movement of goods.[157]

In holding that the obligation was sufficiently precise and unconditional to be relied upon by individuals before national courts, the CJ rejected the argument that the Directive was solely concerned with relations between the MS and the Commission.[158] The result was that CIA could rely on Belgium's failure to comply with its obligations to implement the Directive in its legal action and therefore the Directive had a very

---

[146] ibid 782.
[147] Dougan (n 8) 947.
[148] Dashwood (n 101) 93.
[149] Case C-397-403/01 *Bernhard Pfeiffer and others v Deutsches Rotes Kreuz, Kreisverband Waldshut eV* [2004] ECR *I-08835.*
[150] Marcus Klambert, 'Judicial Implementation of Directives and Anticipatory Indirect Effect: Connecting the Dots' (2006) 43 Common Market Law Review 1251, 1259.
[151] *Pfeiffer* (n 149) *para 118.*
[152] Gerrit Betlem, 'The Doctrine of Consistent Interpretation- Managing Legal Uncertainty' (2002) 22 Oxford Journal of Legal Studies 397, 402.
[153] Craig and De Burca (n 12) 207.
[154] Case C-194/94 *CIA Security International v Signalson and Securitel* [1996] ECR I-2201.
[155] Drake (n 138) 338-339.
[156] Council Directive 83/189/EEC of 28 March 1983 laying down a procedure for the provision of information in the field of technical standards and regulations OJ L109.
[157] Betlem (n 152) 410- 411.
[158] Craig, 'The Legal Effect of Directives: Policy, Rules and Exceptions' (n 122) 364.

marked incidental effect on the action between two private parties.[159] Most notably, as recognised by Dashwood, no attempt was made by the CJ to justify the outcome in the face of the no horizontal direct effect rule which was clearly circumvented. [160]

Dougan noted that *CIA* supports the existence of the primacy model since, rather than employ direct effect, which involves the substitution of domestic law for union law, in *CIA* the CJ excluded domestic law and applied the supremacy principle in order for the directive to create independent effects.[161] However, as has been identified by Craig, if primacy really is the driving force, it is unclear why it should not also demand substitution even in horizontal cases since,[162] in the absence of such substitution, it is through national law that liability arose in the *CIA* case.[163]

Nevertheless, the reality is that, in incidental effect cases, it is the directive that mandates the outcome that constitutes a new legal status quo within the national legal system and thus EU law still takes priority.[164] Moreover, the impact of *CIA* has been extended by its application in the *Unilever*[165] case whereby it applied to civil proceedings arising out of contract.[166] Indeed, *CIA* and *Unilever* verge dangerously close to recognising horizontal effect of directives.[167] Consequently, through *CIA* and *Unilever*, the CJ has imposed EU law upon national courts without a conferral of rights, which clearly advances the argument in favour of the existence of the primacy model in the EU.[168]

3.6 A Future Expansion?

Finally, it is clear that the expansion of this judicial doctrine has not reached saturation point. More recently, there have been suggestions that the Charter of Fundamental Rights[169] may be granted direct effect.[170]

Indeed, as will be discussed in Chapter 3, the CJ has taken to using FR as a tool to expand its power base. In light of this, it is likely that it may use the Charter in a similar way in the future by according it direct effect.

Furthermore, in light of the *Mangold* decision,[171] which has suggested general principles of EU law may now qualify for direct effect, it is possible that FR, as general

---

[159] ibid 365.
[160] Dashwood (n 101) 96.
[161] Dougan (n 6) 952.
[162] Craig, 'The Legal Effect of Directives: Policy, Rules and Exceptions' (n 122) 368.
[163] Betlem (n 152) 412.
[164] Craig, 'The Legal Effect of Directives: Policy, Rules and Exceptions' (n 122) 368.
[165] Case C-443/98 *Unilever Italia SpA v Central Food SpA* [2000] ECR I-7535.
[166] Tridimas 'Black, White, and Shades of Grey: Horizontality of Directives Revisited' (n 141) 345.
[167] ibid.
[168] Prechal (n 79) 1047.
[169] Charter of Fundamental Rights of the European Union (2000/C 364/01).
[170] Eurofound, 'Horizontal Direct Effect' (*Eurofound*) <www.eurofound.europa.eu/areas/industrialrelations/dictionary/definitions/horizontaldirecteffect.htm> accessed on 27 April 2012.
[171] Case C-144/04 *Werner Mangold v Rudiger Helm* [2005] ECR I-9981.

principles of EU law, may gain protection through this route. This would inevitably increase the power holding of the EU and furthers the argument in favour of the primacy model.

3.7 Chapter Summary

Ultimately, whilst the doctrine of direct effect enjoyed a period of continual expansion as a result of judicial activism on part of the CJ, such activism appeared to reach a limit in the case of *Marshall*, whereby the Court announced that directives could not have horizontal direct effect. Whilst this can be recognised as the CJ accepting the limits of its power and heeding the resistance of MS, as might be expected of a court in a relationship of constitutional pluralism, in subsequent case law, the CJ developed numerous exceptions to the *Marshall* rule. In some instances, this has led to what, in practice, amounts to the horizontal effect of directives. This supports the existence of the primacy model. Moreover, Craig has recognised that more exceptions to the *Marshall* rule are likely to arise which would serve to strengthen the argument that the primacy model exists between the EU and MS.[172]

4 FUNDAMENTAL RIGHTS

4.1 Introduction

This Chapter will consider the extent to which the CJ's adoption and subsequent use of FR has assisted in ensuring that EU law is superior to national law in an increasing number of areas, thus supporting the existence of the primacy model between the EU and MS.

Firstly, this Chapter will argue that the CJ introduced FR as general principles of EU law to defend the doctrines of supremacy and direct effect in the face of challenges from the German and Italian Constitutional Courts. Additionally, it will be argued that the CJ has also used FR to increase its market share of competence; first to increase its jurisdictional competence to review MS actions on the basis of FR and also to extend its competence within areas previously regulated by national courts. Despite that there are many instances of such behaviour, this Chapter will consider two instances in particular; regarding MS implementation of, and derogation from, EU law and within the arena of immigration. These case studies have been chosen since, as will be discussed, they have particular repercussions on national sovereignty. Overall, it will be argued that, using FR, the CJ has applied EU rules in areas previously deemed part of national law exclusively. This furthers the contention that EU law is superior to national law and thus supports the existence of the primacy model.

4.2 CJ Activism in Adopting FR as General Principles

Initially, in *Stork*,[173] the CJ refused to allow measures to be challenged on the basis of FR, since there was nothing in the Treaty establishing the European Coal and Steel

[172] Craig, 'The Legal Effect of Directives: Policy, Rules and Exceptions' (n 122) 377.
[173] C-1/58 Frederick Stork & CIE vs. ECSC High Authority of the European Coal and Steel Community [1959] ECR I-17.

Community[174] on which it could draw and since rights recognised in the national constitutions of the MS could not be used to challenge EU law.[175]

Given the reasoning of the CJ in *Stork*, it is clear that it acted outside of its conferred powers years later in the case of *Stauder* when it affirmed the recognition of FR as general principles of EU law.[176] Whilst some argue that the recognition of FR in *Stauder* was intended as a concession to national courts, it will be argued that the adoption of FR was a measure taken by the CJ in order to protect the direct effect and supremacy doctrines. Evidence of such activism demonstrates the shift away from the traditional model and supports that the relationship between the EU and MS resembles the primacy model.

The context in which *Stauder* was decided is important. The case law of the German and Italian Constitutional Courts in the 1960's was characterised by a fear that FR would be eroded as the competences of the EU increased.[177] Indeed, in 1973 in the *Frontini* Case,[178] the Constitutional Court of Italy reserved the right to declare the treaty incompatible with the constitution in the 'unlikely' event of EU legislation being made which violated FR.[179] Similarly, in 1974 in *Internationale Handelsgesellschaft*[180] the GCC held that, so long as the EU lacked a codified catalogue of FR, the guarantee of FR in the German constitution would prevail over EU law.[181]

This has led some commentators, such as Albi, to contend that the introduction of FR as general principles was a concession by the CJ to national courts.[182] Albi argues that 'cooperative constitutionalism'[183] exists between the CJ and national constitutional courts. This model of cooperative constitutionalism mirrors that of Barents' model of constitutional pluralism. Indeed, Albi and Barents agree that the adoption of FR supports the existence of a heterarchial structure in which legal orders adapt themselves autonomously to legal facts and developments in the other legal order[184] just as the CJ autonomously incorporated protection of FR into EU law to avoid the danger that national constitutional courts would declare EU rules inapplicable in their national territory.[185]

---

[174] Treaty establishing European Coal and Steel Community, signed on 18 April 1951.
[175] Francis G Jacobs, 'Human Rights in the European Union: the Role of the Court of Justice' (2001) 26 European Law Review 331, 332.
[176] C-29/69 Eric Stauder v City of Ulm [1969] ECR I-419.
[177] Jason Coppel and Aidan O'Neill, 'The European Court of Justice: Taking Rights Seriously?' (1992) 12 Legal Studies 227, 228.
[178] Frontini v. Ministero delle Finanze case 183 of 27/12/73 reported in [1974] 2 CMLR 383-90.
[179] MH Mendelson, 'The European Court of Justice and Human Rights' (1981) Yearbook of European Law 125, 131.
[180] Internationale Handelsgesellschaft v Einfuhr und Vorratselle fur Getreide und Futtermittel (n 35).
[181] ibid.
[182] Albi (n 49), 30.
[183] ibid 27.
[184] Barents (n 7) 441.
[185] ibid.

However, as observed by De Burca and Aschenbrenner, an alternative reading of the case law leads to the conclusion that the CJ discovered FR as general principles of EU law at least partly in response to the reluctance of MS to accept fully the principles of direct effect and supremacy.[186] Indeed, when considering the importance of the doctrines of supremacy and direct effect portrayed in Chapters One and Two, De Burca and Aschenbrenner's argument seems most persuasive given the threat[187] that the case law from Italy and Germany posed to both the principles. This challenges the idea that the introduction of FR supports the existence of constitutional pluralism.

As identified, the CJ acted outside of its conferred powers in *Stauder* and therefore the case demonstrates that there has been a shift away from the traditional model. Furthermore, the adoption of FR as general principles of EU law by the CJ, as will be seen throughout the remainder of the Chapter, has led to an increasing number of issues being decided through EU law rather than national law which undoubtedly strengthens the argument in favour of the primacy model.

4.3 An Offensive Use of FR

There is much evidence to indicate that, since the recognition of FR as general principles of EU law, the CJ has continued to use FR to expand its competences and, ultimately, to increase the number of decisions which are decided on the basis of EU law. Coppel and O'Neil have characterised such use of FR as 'offensive'[188] and identify two ways in which the CJ has used FR offensively. Firstly, the CJ has undertaken an expansion of its jurisdiction regarding when it can review MS action for compliance with FR. Secondly, the CJ has extended the use of FR in areas of EU law previously untouched by such concepts;[189] notably with regard to MS derogation from EU law and immigration law.

4.3.1 Jurisdictional Expansion

Initially, the CJ considered the issue of reviewing MS action on grounds of FR in *Cinetheque*[190] which involved a piece of French legislation prohibiting the marketing of video cassettes of a film for a period of one year following the showing of the film in cinemas.[191] This legislation was challenged on the ground that it violated the principle of free expression enshrined in Article 10 of the European Convention on Human Rights (ECHR).[192]

---

[186] Grainne De Burca and Jo Beatrix Aschenbrenner, 'The Development of European Constitutionalism and the Role of Fundamental Rights' (2002-2003) 9 Columbia Journal of European Law 355, 368.
[187] Mendelson (n 179) 132.
[188] Coppel and O'Neill (n 177) 227.
[189] ibid.
[190] C-60-61/84 Cinetheque v Federation nationale des cinemas francais [1985] ECR I-2605.
[191] Darcy S Binder, 'The European Court of Justice and the Protection of Fundamental Rights in the European Community: New Developments and Future Possibilities in Expanding Fundamental Rights Review to Member State Action' (Jean Monet Center for International and Regional Economic Law and Justice) <http://centers.law.nyu.edu/jeanmonnet/archive/papers/95/9504ind.html> accessed on 27 April 2012.
[192] ibid.

When determining the scope of its jurisdiction to review MS action on the basis of FR protection, the CJ stated in *Cinetheque* that it had no power to examine compatibility with the ECHR of national legislation that concerns an area that falls within the jurisdiction of the national legislature.[193] Indeed, there are some areas of law that fall within the scope of both EU law and national law. In *Cinetheque*, the CJ appears to have been conscious of its jurisdiction within these areas.[194] This is clear since areas of shared competence would fall within the jurisdiction of the national legislator and thus, according to *Cinetheque*, would not be reviewed on FR grounds by the CJ. Such a consciousness might be expected of courts working together in constitutional pluralism.

However, the stance of the CJ seemed to change in *Demirel*;[195] this involved a course of action for the annulment of an order to expel Mrs Demirel, a pregnant Turkish worker, from Germany where she had been residing on a tourist visa with her husband. In its preliminary ruling, the CJ had to address arguments put forward regarding Article 8 ECHR. In addressing this issue, the Court subtly changed[196] the jurisdictional test by stating it has no power to examine the compatibility with the ECHR of national legislation lying outside the scope of EU law.[197] Coppel and O'Neill have argued that such a nuance has revolutionised the impact of FR considerations on national administrative and legislative action.[198] Indeed, in contrast to *Cinetheque*, the jurisdictional test in *Demirel* appears to allow an increase in the CJ's invasiveness in areas of shared competence, which is uncharacteristic of constitutional pluralism.

Following *Demirel*, Coppel and O'Neil have argued that the only MS actions that the CJ might decline to vet on FR grounds are those that occur in an area of exclusive MS jurisdiction.[199] However, this has been criticised by Weiler and Lockhart who state that such a view is 'simply wrong'.[200] Weiler and Lockhart justify this by reference to the outcome of the *Demirel* case and point out that, although *Demirel* clearly involved a matter of shared competence, the CJ declined to exercise FR jurisdiction.[201] Whilst *Demirel* does indicate that, in practice, not everything outside the exclusive jurisdiction of the MS would be reviewed,[202] it does not negate that there has been an expansion of jurisdiction from *Cinetheque* to *Demirel*. Indeed, even Weiler and Lockhart do not contest the point that *Demirel* uses a wide formula[203] and, despite the outcome in

[193] Cinetheque (n 190) para 26.
[194] Pescatore, 'La Cour de Justice des Communautes Europeens et la Convetion Euroeene des Droits de l'Homme' in Protecting Human Rights: The European Dimension (1988) 441, 446 cited in Coppel and O'Neill, (n 177) 235.
[195] C-12/86 Maryem Demirel v Stadt Schwabisch Gmund [1987] ECR I-3719.
[196] Coppel and O'Neill (n 177) 235.
[197]Demirel (n 195) Para 28.
[198] Coppel and O'Neill (n 177) 235.
[199] ibid 236.
[200] JHH Weiler and Nicolas JS Lockhart, '"Taking Rights Seriously" Seriously: The European Court and its Fundamental Rights Jurisprudence- Part 1' (1995) 32 Common Market Law Review 51, 62.
[201] ibid 63.
[202] ibid.
[203] ibid.

Demirel, this surely has potentially far-reaching consequences, as will be seen in Section 4.3.1.1.

These cases have increased the ability of the CJ to review national authorities for compliance with FR, thus increasing the CJ's jurisdiction within areas of shared competence. Since this is likely to result in an increasing number of decisions made through EU, as opposed to national law, this furthers the contention that the primacy model exists between the EU and MS.

4.3.1.1 CJ Invading the Areas of MS Implementation of, and Derogation from, EU Law Armed with FR

The expansion of the CJ's jurisdiction in *Demirel* increased the potential invasiveness of the Court in areas of shared competence. In this way, *Demirel* can be seen to have paved the way[204] for *Wachauf*[205] and *ERT*,[206] both of which involved areas of shared competence. In these cases, the CJ announced a right of review for both MS implementation of EU law and MS derogations from EU law on FR grounds, despite that these would typically be decisions reserved for the national courts. Therefore, this will lead to more issues being resolved according to EU law, not national law, which supports the existence of the primacy model.

In *Wachauf*, for the first time, the CJ determined that MS were required to observe FR when implementing EU law, despite that it had previously been thought that such an obligation only applied to EU institutions.[207] On the one hand, it has been argued as self evident that MS should be bound as a matter of EU law to protect FR, [208] since MS often act as the executive branch of the EU.[209] Additionally, it has been argued that to leave this review to the national constitutions would endanger the uniformity of EU law.[210] However, as addressed in Section 2.2, uniformity is an insufficient justification for the CJ to extend its competences, and so the strength of this argument can be questioned. Moreover, for the purpose of this thesis, the question should not be whether MS actions should be subject to FR review, but whether *Wacauf* represents an extension of the CJ's ability to rule over an area of national law, which it clearly does.

Furthermore, in *ERT*, the CJ interpreted the effect of *Wachauf* in very broad terms[211] and declared it had a duty to ensure that MS adequately respect FR, as part of EU law, when adopting measures derogating from EU law.[212] As AG Jacobs has noted, once a

[204] Coppel and O'Neil (n 177) 235.
[205] C-5/88 ***Hubert Wachauf v Bundesamt für Ernährung und Forstwirtschaft*** *[1989] ECR I-2609.*
[206] C-260/89 *Elliniki Radiophonia Tileorassi AE (ERT) v Dimotiki Etairia Pliroforissis and Sotirios Kouvelas [1991] ECR I-02925.*
[207] Mendelson (n 179) 151.
[208] Jacobs (n 175) 334.
[209] Weiler and Lockhart (n 200) 74.
[210] Jacobs (n 175) 333.
[211] Coppel and O'Neil (n 177) 233.
[212] Craig and de Burca, (n 12) 386.

derogating act is justified from the perspective of EU law, the act might still infringe FR; however this would normally be a matter for national law to determine and not EU law.[213] Thus, since the decision in ERT means that this is now also a matter for the CJ to decide on, the CJ can clearly be seen to have used FR as a tool to increase the number of decisions based on EU law rather than national law.

Weiler and Lockhart have attempted to justify the decision in *ERT* by arguing that the scope of the derogation, and the conditions for its employment, are all creatures of EU law.[214] They further this point by inviting their readers to imagine the state of the common market if each MS could determine, by reference to its own laws and values, (without reference to EU law) what was or was not covered by the prohibition and its derogation.[215]

However, for the purposes of this thesis, the most salient question is not whether the CJ should have jurisdiction to examine MS derogations on FR grounds, but whether they in fact do have such jurisdiction conferred upon them by the treaty to justify their actions in *ERT*. Indeed, the evidence would suggest they do not. This conclusion can be supported by reference to the Charter of FR which,[216] since the Treaty of Lisbon,[217] takes legal effect. Whilst its provisions are expressly addressed to MS when implementing EU law, they are not expressed to be applicable to MS when derogating from the treaty.[218] Moreover, the suggestion that this is an, 'inadvertent omission' seems doubtful.[219]

Thus, as a result of *Demirel* and through *Wachauf* and *ERT*, the CJ can be seen to have used FR to extend its jurisdiction into areas that were previously matters for national courts and thus apply EU rules in preference to national law. The extent of this is most blatant in the context of *ERT*, since derogations from EU law are bound up with fundamental notions governing the relationship between states and their citizens, and for this reason one cannot fail to appreciate the potential effect of the *ERT* judgement on national sovereignty.[220] The ability of the CJ to claim priority for EU law across areas of law proximate to national sovereignty adds weight to the contention that the power sharing arrangement between the EU and MS resembles the primacy model.

4.3.2 CJ Invading the Area of Immigration Law Armed with FR

Furthermore, in the area of immigration law, the CJ has claimed superiority for EU law over scenarios previously considered purely domestic situations. As Lawson has identified, family reunification in particular is generally believed to account for a large

[213] Jacobs (176) 336.
[214] Weiler and Lockhart (200) 76.
[215] ibid.
[216] Charter of Fundamental Rights of the European Union (n 169).
[217] (n 65).
[218] Jacobs (n 175) 338.
[219] ibid 338-339.
[220] G Federico Mancini and David T Keeling, 'From CILFIT to ERT: the Constitutional Challenge Facing the European Court' (1991) Yearbook of European Law 1, 2.

proportion of immigration[221] and it is within this context that the CJ can most blatantly be seen to have extended its jurisdiction. This development is inherently controversial since family reunification is a major political issue at the MS level.[222]

In *Akrich*,[223] the CJ determined that, for third country national family members of an EU national to enjoy family reunification rights in one MS, they must have been lawfully resident in another EU MS prior to their application under the EU rules (the prior lawful residence rule). Third country nationals living outside of the EU could thus not rely on EU free movement to make their first entry into the Union - an issue of external borders and national immigration policy. Indeed, as identified by AG Geelhoed, this aspect of the ruling accords fully with the division of competences between the EU and MS[224] since the reach of EU law is limited and a situation needs to fall within the scope of the treaties for EU law to be triggered.[225] Thus, in relation to the free movement provisions, absent movement between two EU countries (a cross border element) as in *Akrich*, the treaty is not triggered and the situation is a wholly internal one to be regulated by national law.[226]

However, the CJ chose not to apply *Akrich* in the case of *Jia*.[227] Rather than insist on a cross border element to trigger the application of the treaties, the mere fact that someone satisfied the familial link with a migrant economic actor in combination with the fact that there was some kind of interstate movement was deemed sufficient for granting family reunification rights under EU law.[228] The CJ justified this departure by referring to the facts of *Akrich*; Mr Akrich, a non EU resident and his wife married in Ireland, where Mr Akrich had been deported to from the UK, with the intention to take advantage of EU law so that Mr Akrich could then gain a right to reside in the UK whilst avoiding UK immigration laws. Thus, the CJ in *Jia* concluded that the outcome of *Akrich* was tailored 'in order to combat steps taken by members of the family of an EU national who did not meet the conditions laid down by national law for entry and residence in a MS'.[229] This allowed the CJ to conclude 'that the condition of previous lawful residence in another MS, as formulated in the judgment in Akrich, cannot be transposed to the present case and thus cannot apply to such a situation.'[230]

---

[221]Rick Lawson, 'Case Comment: Court of Justice of the European Communities: Family Reunification and the Union's Charter of Fundamental Rights, Judgment of 27 June 2006, Case C-540/03, Parliament v Council' (2007) 3 European Constitutional Law Review 324.
[222] ibid.
[223] C-109/01 *Secretary of State for the Home Department v Akrich* [2003] ECR I-9607.
[224] Alina Tryfonidou, 'Jia or "Carpenter II": the Edge of Reason' (2007) 32 European Law Review 908, 911.
[225] Alicia Hinarejos, 'Case Comment: Extending Citizenship and the Scope of EU Law' (2011) 70 Cambridge Law Journal 309.
[226] ibid 310.
[227] C-1/05 *Yunying Jia v Migrationsverket* [2007] Q.B. 545.
[228]Tryfonidou 'Jia or "Carpenter II": the Edge of Reason' (n 224) 914.
[229] *Jia* (n 227) para 30.
[230] *Jia* (n 227) para 32.

However, whilst this might be seen as a justification for not applying the prior lawful residence rule, this does not explain why the CJ adopted such a flexible approach in finding a link with EU law. Indeed, as stated by Tryfonidou, the main characteristic of the *Jia* case is that the search for an equitable solution in the individual case appeared to be of higher importance to the CJ, than preserving a coherent approach respecting cardinal principles of EU law such as the principle of attributed competence.[231] The interpretation of the CJ in *Jia* as searching for an equitable solution suggests that the Court might have been taking FR considerations into account. Indeed, whilst, in the case of *Jia*, the CJ avoided laying out its justification for this flexible approach, an analysis of subsequent case law suggests that FR influenced the CJ and enabled it to claim superiority for EU law in immigration cases.

In its *Metock* judgment, the CJ went even further in expanding its jurisdiction.[232] Firstly, it contended that the application of the prior lawful residence rule expounded in *Akrich* was in violation of the Directive 2004/38.[233][234] Controversially, the CJ determined that third country nationals benefiting from family reunification rights may come into any MS in the EU from anywhere in the world. Through *Metock*, the CJ has assumed competence over national immigration powers on the basis that Directive 2004/38/EC[235] gave the EU competence to enact the necessary measures to bring about freedom of movement for Union citizens.[236] It then went on to state that 'if Union citizens were not allowed to lead a normal family life in the host MS, the exercise of the freedoms they are guaranteed by the Treaty would be seriously obstructed.'[237] Thus, the CJ in *Metock* clearly relied on FR; namely that of the right to a private and family life, to justify the extension of its jurisdiction to cover immigration cases lacking a cross boarder element, which was previously required to trigger the application of EU law. Tryfonidou defends this reasoning by the Court, stating it is an important and admirable step to be taken.[238] However, whether this is true or not is irrelevant to this thesis. What is most salient is that this was clearly an instance whereby the CJ used FR as a tool to avoid the parameters of its competence and expand the number of decisions based on EU law rather than national law. Therefore, whilst *Akrich* was identified as according with the division of competences between EU and MS, this can clearly not be said for *Metock*.

Furthermore, after *Akrich*, Tryfonidou predicted that, had the CJ in *Akrich* found in favour of Mr Akrich, the MS would likely have vehemently objected to such an unwarranted intrusion into an area as central to their sovereignty as that of immigration

[231] Tryfonidou 'Jia or "Carpenter II": the Edge of Reason' (n 224) 914.
[232] *Metock* (n 53).
[233] Alina Tryfonidou, 'Family Reunification Rights of (Migrant) Union Citizens: Towards a More Liberal Approach' (2009) 15 European Law Journal 634, 645.
[234] (n 54).
[235] ibid.
[236] *Metock* (n 54) para 61.
[237] *Metock* (n 54) para 62.
[238] Tryfonidou 'Jia or "Carpenter II": the Edge of Reason' (n 224) 917.

control.[239] However, as discussed in Chapter 2, the objection of MS during the case of *Metock* was unsuccessful and no objection materialised subsequent to the ruling. Such lack of resistance challenges the existence of constitutional pluralism in Europe.

Ultimately, by using FR to justify overturning *Akrich* and dilute the requirement of a cross border element, the CJ paved the way for the supremacy principle to then ensure EU law emerged as supreme in immigration cases regarding family reunification (as was discussed in Section 2.4.2)

As the EU is becoming increasingly superior to national law in areas of national sensitivity, thus lends weight to the argument that the primacy model reflects the power sharing arrangement between the EU and MS.

4.4 Protocol on the Application of the Charter of FR

On a final note, it might be argued that the extent to which the CJ can proceed down the road of expanding the EU's power through FR is questionable in light of the Protocol on the application of the Charter of FR whereby Britain, Poland and the Czech Republic have attempted to limit the applicability of the Charter within their countries.[240] The Protocol states 'nothing in the charter creates justiciable rights' applicable to those countries.[241] However, as identified by Craig and De Burca, the Protocol does not propose to overturn the prior case law of the CJ where it claimed the competence to review MS actions, for example, in light of FR as general principles of the EU.[242] In light of this, the Protocol is thought to be largely declaratory and cannot therefore be seen as a hurdle to the further expansion of the CJ's jurisdiction with regard to FR.[243]

4.5 Chapter Summary

In conclusion, this Chapter has argued that the CJ has used FR as a mechanism to increase the number of decisions made through EU law as opposed to national law. Initially, it was argued that the CJ used FR in order to protect the principles of supremacy and direct effect that, as identified previously in this thesis, are the primary tools through which the CJ has increased the power of the EU over the MS. However, the CJ has also been demonstrated to have used FR in an offensive way which has been identified in three circumstances. Firstly, the CJ expanded the scope of its jurisdiction to review MS actions on the basis of FR. Secondly, through this expansion the CJ claimed competence to review MS action when implementing EU law and derogating from EU law on the basis of FR. Finally, and most controversially, the CJ used FR in the context of immigration law to the effect that national immigration laws are circumvented in circumstances of family reunification. The culmination of the above case law has seen a substantial shift in the number of decisions being decided through

---

[239] ibid 914.
[240] Protocol on the application of the Charter of Fundamental Rights of the European Union to Poland and to the United Kingdom, Lisbon Treaty (n 65).
[241] ibid Article 1(2).
[242] Craig and De Burca (n 12) 395.
[243] ibid.

EU law, even in areas of national sensitivity. Moreover, since the CJ can be seen to have acted outside of its conferred powers through the treaties, such action lends force to the argument that the power sharing arrangement has developed away from the traditional model, toward the primacy model.

5 CONCLUSION

This article sought to argue that through CJ activism, the EU's power sharing arrangement has shifted away from the traditional model toward the primacy model. Indeed, the ability of the CJ to act outside of its conferred powers and increase the number of occasions whereby EU law supersedes national law is characteristic of this power structure.

This move toward the primacy model has been demonstrated throughout this thesis by focussing on three areas where the shift of power to the EU is most evident. However, it is acknowledged that there are other instances where the Court can be deemed to have applied EU law in area previously within the remit of national law.

Firstly, the development and expansion of the principle of supremacy was considered. It was observed that the introduction of this principle saw the CJ act outside its conferred powers since there was a notable lack of evidence from the treaties and from national constitutions that the introduction of the doctrine was intended by the MS. Furthermore, it was noted that some academics nevertheless argued that the principle of supremacy accorded with constitutional pluralism, since the doctrine was bi-dimensional and it was, therefore, within the power of the MS to regulate the use of the doctrine. However, as was demonstrated throughout the remainder of Chapter 2, the doctrine of supremacy has become increasingly one-dimensional. Indeed, it was observed that the MS themselves actually reduced the circumstances in which they had the ability to regulate the use of the doctrine. Additionally, and most importantly for the purpose of this thesis, it was also demonstrated that even where such regulation was attempted by MS, such as in the case of *Metock*,[244] this was unsuccessful and the CJ imposed supremacy on MS nonetheless. This challenges Barents' model of constitutional pluralism, whereby legal orders are meant to resist each other's advances successfully. Moreover, since the second dimension of the doctrine has been eliminated in practice, the CJ is the sole regulator of the doctrine that undoubtedly places a lot of power in the hands of the EU. Indeed, the Chapter concludes that EU law has, through the doctrine of supremacy, emerged as hierarchically supreme and the contention that the CJ has led the relationship between the EU and MS toward the primacy model is furthered.

Secondly, the doctrine of direct effect was considered. Primarily, it was identified that the resistance by MS to the introduction of the doctrine in *Van Gend en Loos* was unsuccessful.[245] However, the subsequent behaviour of the CJ in *Marshall* called into

[244] *Metock* (n 54).
[245] *Van Gend en Loos* (n 80).

question whether the CJ is really guiding the EU and the MS toward the primacy model,[246] since it refused to recognise the horizontal direct effect of directives.

It was observed that the most convincing reason for this refusal was that it was a result of the initial resistance toward the vertical direct effect of directives and thus it seemed that the CJ had finally acknowledged its limits and accepted that it was in a relationship of constitutional pluralism. Nevertheless, over time, it was addressed that the CJ introduced three key areas of case law that stand in contradiction to the *Marshall* ruling. In light of these exceptions, it was identified that, in practice, the CJ has accorded direct effect to directives in horizontal situations. As a result, there are an increasing number of situations where EU law applies even when the MS have not implemented directives. It was argued that this challenges the existence of constitutional pluralism and, rather, furthers the argument in favour of the move toward the primacy model between the EU and MS.

Thirdly, the development of FR was addressed. The argument that FR were developed by the CJ as a concession to the MS was challenged and it was argued that FR were, instead, developed in order to defend the doctrines of supremacy and direct effect; and thus to ensure the hierarchical primacy of EU law. Furthermore, it was identified that the CJ subsequently employed FR to increase the application of EU law in areas of sensitivity such as MS derogation and immigration law, previously deemed part of national law exclusively. The latter aspect of the CJs use of FR is the most controversial in light of the repercussions on national sovereignty. The ability of the Court to claim competence in such sensitive areas without successful resistance from MS challenges the idea that the EU and MS are in a relationship of constitutional pluralism. Rather, it suggests that the primacy model exists.

Finally, it has been identified that this development toward the primacy model is likely to continue since more exceptions to the *Marshall* rule are likely to arise and it has been suggested that the CJ may find the Charter of Fundamental Rights[247] is directly effective. In conclusion, the culmination of the development of supremacy, direct effect and FR discredits Barents' theory that constitutional pluralism exists between the MS and the EU.

Instead, it has been demonstrated that CJ activism in the development of these three doctrines, which have been largely accepted by MS, has ensured that EU law is in every way capable of superseding national law. This is demonstrated since supremacy ensures priority for EU law, direct effect ensures the application of EU law in the case of treaty articles, regulations, decisions and directives (regardless of MS implementation) and, through FR, the CJ has increasingly decided issues through EU law which were previously considered part of national law exclusively. That EU law has developed such

[246] *Marshall* (n 115).
[247] (n 168).

a superior role supports that the EU and MS are in a governmental power sharing arrangement represented by the primacy model.

A CLOSE LOOK ON IMPROVING DEMOCRACY IN THE EUROPEAN UNION

ESBEN POULSEN[*]

The European Union as a multi-national institution has an important role in influencing the policies of its Member States. The question is what measures should be taken in order to improve the democratic inputs in European Union. Therefore, it is examined whether a parliamentary system in the EU that imitates the national political systems would be effective, and whether this change would resolve its democratic deficits. Even though the adoption of parliamentary standards of 'deliberative democracy' would cause structural changes in the EU and change its identity, it would minimise the democratic deficits but not solve them completely.

Paul Craig's analysis of the Lisbon Treaty proclaims that a move towards a 'more truly 'parliamentary' system has improved democratic input.[1] He could be suggesting that embracing parliamentarianism can help bridge the 'gap' between electors and the European Union (hereinafter EU). Yet, he gloomily asserts that there are '*structural reasons* why, although there can be improvements in this respect, [democratic input in EU decision making] *cannot be perfectly realized*'.[2] Is the solution to the democratic deficit problem in the EU then an elaborate parliamentary democracy? Answering in the affirmative to this question is problematic because the structure of the EU does not marry well with the parliamentary notion of democratic input. *Should* the democratic deficit 'problem' be solved by mimicking national parliaments, if it is not possible identifying the root of the problem?

This essay will attempt to explain that the structure of the EU prohibits parliamentary antidotes. It is suggested that there is justification that signals the narrowing, but not solving, of the democratic deficit through parliamentary means because European Union democracy 'cannot, and need not, be shaped in analogy to that of a State'.[3] Thus, on the metaphysical level, but also as a matter of practical politics, the structural identity of the EU defeats the normative question of statehood analogisms. Parliamentary remedies may help narrow the democratic deficit, but they are insufficient because the structure of the EU does not embody that of a parliamentary state.

---

[*] Esben Poulsen, Newcastle University, Law LLB State Two.

[1] Paul Craig, *The Lisbon Treaty: Law, Politics, and Treaty Reform* (OUP 2010) 72.

[2] Craig (n 1) (emphasis added).

[3] Daniel Thym, 'In the Name of Sovereign Statehood: a Critical Introduction to the Lisbon Judgment of the German Constitutional Court' [2009] 46 CML Rev 1795, 1813.

The concept of a democratic deficit—or perhaps democratic illegitimacy—is often held against a conceptually 'frankly utopian' mirror of 'deliberative democracy'.[4] Andrew Moravcsik defends democratic illegitimacy in the EU by invoking 'relaxed' standards. More specifically, the EU is a 'multinational body' that lacks common heritage (in various forms).[5] What justifies the EU as a 'super-majoritarian' institution, he concludes, are 'robust' means of accountability.[6] These could occur first, directly by European elections to the European Parliament and second, indirectly through elections of national officials.[7] If these apologetic standards are presumed, the EU should not structurally change—it is already democratically accountable. The EU may not be expected to be an 'ancient, Westminster-style' type of governmental body.[8] To conclude from Moravcsik's article on a general level, the EU cannot—structurally—be expected to take on the aspect of a national state, largely because that conflicts with its purpose. These arguments remarkably predate the Lisbon Treaty, yet they remain appropriate,[9] as the functioning of the EU is a 'representative democracy', directly and indirectly accountable as explained above.[10]

Moravcsik presents an attractive proposition; approaching the EU as a functional institution rather than a parliamentary institution is appealing. He is setting a 'fair standard' and considers the practical implications of a, now twenty-seven, Member State cooperative.[11] That is to say, the EU's purposes override its imperfections. The EU has perhaps not assumed the role of a super-state. Instead, a neofunctionalist process of European integration has meant that the cooperative has '[spilt] over' from 'non-controversial, technical' sectors to those of a 'greater political salience'.[12] The cooperative has drained power from national parliaments but has not assumed their role.[13] Layman and lawyer alike will be uneasy towards that point. If the EU has greater, perhaps parliamentary, powers, but remains a cooperative, it ought to narrow the democratic deficit. Democratic legitimacy is then a serious concern. The democratic deficit must be limited because of the EU's parliamentary-like functions, yet its structure makes it unlikely to be solved completely. In fact, it is insufficient to justify analogies to that of a state for that reason: the purpose and structure of the EU are not to embody a 'Westminster-style (…) form of deliberative democracy'.[14] It is not to say

[4] Andrew Moravcsik, 'In Defence of the 'Democratic Deficit': Reassessing Legitimacy in the European Union' [2002] 40 JCMS 603, 605.
[5] Moravcsik (n 4) 604.
[6] Moravcsik (n 4) 603, 611.
[7] Moravcsik (n 4) 611.
[8] Moravcsik (n 4).
[9] Treaty of Lisbon amending the Treaty on European Union and the Treaty establishing the European Community [2007] OJ C306/01.
[10] Consolidated Versions of the Treaty on European Union and the Treaty on the Functioning of the European Union [2010] OJ C83/01, TEU, art 10; Moravcsik (n 7).
[11] TEU (n 10) art 1.
[12] Juliet Lodge, *The European Community and the Challenge of the Future* (Juliet Lodge ed, Pinter 1993), xix.
[13] Andrew Follesdal and Simon Hix, 'Why There is a Democratic Deficit in the EU: A Response to Majone and Moravcsik' [2006] 44 JCMS 533, 534-35.
[14] Moravcsik (n 4).

that political, psychological or statistical concerns of the citizens' influence on the EU are misplaced. It is to say that such concerns, while perhaps genuine, are unjustified if one appreciates the structural identity of the EU.

The role of the democratic deficit in the EU is usefully explained in Daniel Thym's case note of the 'Lisbon Decision' of the German *Bundesverfassungsgericht* (or Constitutional Court). The court examined *inter alia* the question of European democracy and its legitimacy.[15] The structural boundaries and identity of the EU are also elaborated on. It had previously said that the critical factor was that 'the democratic bases of the European Union' should be 'built-up in step with integration'.[16] This connects with the neofunctionalist perspective of EU's development, namely to further 'spill over' into other sectors demands further legitimacy.[17] This is an analogy of logical progression. Structural development of the EU entails a commitment to further narrow the democratic deficit. Its Lisbon judgment asserts '[the EU] complies with democratic principles because a qualitative look at the structure of its responsibility and of its [government] reveals that it is exactly not laid out in analogy to a state.'[18] That is to say, the foundation and current structure of the EU are democratically justifiable. The key reason is that the EU does not assume the structural identity of a state. This lends support to the analysis of the importance of structuralism above. It is not to accept apologetic standards, but rather that the structure may mean parliamentary remedies are to some extent insufficient or inappropriate.

In Thym's interpretation the court states 'the absence of an independent constituent power at European level hinders the direct democratic legitimacy of European politics'.[19] In other words, because the EU's structure is different to that of a state, its politics cannot be directly democratically legitimate. That is a hard blow to structuralism; it is getting to the heart of the problem by accepting the structuralist argument. The EU's powers, which affect citizens of the EU, *cannot* be directly accountable to them. Yet the 'matter of fact' analysis above illustrates the need to accept exactly that. Direct accountability was not the purpose of the EU and that is evident in its structural identity. To think otherwise is to analogise to the structural identity of a State. That precisely reveals the flaws of applying parliamentary mechanisms to the EU. Granted, it may increase democratic input. However, it cannot be the *absolute* solution, given the purpose of the EU to establish a cooperative 'to attain objectives [the High Contracting Parties] have in common'.[20]

Parliamentary mechanisms can, thus, increase democratic input, but fall short of completely remedying the democratic deficit problem. If it were the complete solution,

[15] Thym (n 3) 1812.
[16] Thym (n 3).
[17] Lodge (n 12).
[18] Thym (n 3).
[19] Thym (n 3) 1819.
[20] TEU (n 10) art 1.

it follows that the EU would reform into statehood. Thym's analysis concludes that further 'autonomy of European decision-making by means of independent government structures with a powerful parliament would establish European statehood'.[21] The court's concern here was, by and large, that a truly and completely parliamentary system in the EU would establish a federal union.[22] This would solve the democratic deficit problem, but change the purpose of the EU. It would be a remarkable change of balance vis-à-vis national parliaments' sovereignty.[23] That may also be the explanation behind the peculiar choice of words in the treaties, namely 'an ever closer union' but not an absolute union.[24] The purpose and identity of the EU make it resistant to increasing democratic input by parliamentary means because this relies heavily on the national state analogy that the EU dares not to embrace. Thus, the paradox is that embracing direct democratic legitimacy is *contra* the aims of the EU.

Moravcsik's argument is that the EU is already directly and indirectly accountable. The EU 'need not evolve towards State analogy'.[25] However, this does not necessarily dismiss concerns of a gap between elector and the EU. On this, the *Bundesverfassungsgericht* emphasised the responsibility of the Member States to 'narrowing the democratic deficit', while welcoming means of 'transparent and participatory decision-making'.[26] Craig suggested that increasing democratic input could be done by parliamentary means. In turn, the German *Bundesverfassungsgericht* was concerned with generally narrowing the democratic deficit without moving towards an analogy of a state. Thym notes the court's appreciation of 'novel forms' of involving the electors in the internal decision-making process.[27] This would support the seemingly paradoxical side of the structuralist argument above.

The Lisbon Treaty introduced an emphasis on participatory democracy.[28] *Inter alia*, there is a clear emphasis on an 'open, transparent and regular dialogue' with society as a whole.[29] This is embodied in the particularly striking concept of the 'Citizens' Initiative' (hereinafter CI). At its conceptual stage, it provides 'not less than one million citizens' of the EU with the 'power' to bring a matter for legislation before the European Commission. As a practical matter, it is considerably illusory. Given the procedures, set by the European Parliament and Council, are met, how compelling is a CI's proposal? There is no empirical evidence on this, as the CI's procedures and conditions apply from 1 April 2012.[30] At its worst, a CI will perhaps only be a subject matter on

[21] Thym (n 3).
[22] Thym (n 3) 1821.
[23] Thym (n 3) 1819, 1821.
[24] TEU (n 10) art 1.
[25] Thym (n 3) 1815.
[26] Thym (n 3) 1813, 1820.
[27] Thym (n 3) 1816.
[28] Kluwer Law International, 'Editorial comments: Direct democracy and the European Union… is that a threat or a promise?' [2008] 45 CML Rev 929.
[29] TEU (n 10) art 11(2).
[30] Regulation (EU) No 211/2011 of the European Parliament and of the Council of 16 February 2011 on the citizens' initiative [2011] OJ L65/1, art 23.

the parliamentary agenda.[31] The numerous requirements that follow also do not enhance the prospect of a powerful, successful application of the concept, as for example the eligible signatories must be from 'at least one quarter of all Member States'.[32] There are also practical concerns as to the administration of such an endeavour. For example, it must be considered whether the verification of individual information for each of the one million citizens is administratively workable. There is here a question of resources and a 'serious thought' about overcoming the 'inevitable logistical challenges'.[33] However, it remains a conceptually intriguing work-around of the issue of structuralism and parliamentary democracy.

Relating to Craig's point above, it allows for improved democratic input, yet is a significantly different option from that of a state-like parliament. It is powerfully democratic to pursue the goal of having citizens set the legislative agenda through the CI mechanism even though the logistical implications make it stand out as a cumbersome means to improve democratic input.

The strengths of the structuralist argument is that it has due regard to the functional, if not neofunctional, identity of the EU. This is evident in the *Bundesverfassungs-gericht* decision, according to which the EU 'cannot and need not' to be shaped as a state.[34] This defeats its purpose and would challenge national parliaments' sovereignty.[35] If the EU fully embraced parliamentarianism, the democratic deficit would be solved but it would constitute European statehood such as establishing a type of federal union. Even though this attempt would redefine the EU as an absolute union, yet it is evident that brings challenges to its constitutionality and is *contra* its purpose of a non-absolute cooperative.

To conclude, the above analysis of structuralism shows that there is justification for narrowing the democratic deficit. However, if it is attempted to solve it completely by mimicking national parliaments, it entails a new structural identity of the EU. Simply adopting parliamentary standards of 'deliberative democracy' would, with due regard to the structuralist argument, be approaching a different state-like identity.[36] It follows that such changes are insufficient to remedy the democratic deficit in the EU.

On the other hand, the 'case study' of the CI shows that, while perhaps not practically perfect, the EU is committed to increase democratic input by participatory democracy. Thus, it narrows, but does not necessarily solve, the democratic deficit. That is not to say the gap between the elector and the EU does not embrace more than what is described above. However, this is not an essay in statistics or politics asking, 'what

[31] Editorial comments (n 27) 937.
[32] Regulation on the citizens' initiative (n 29) art 2(1).
[33] Editorial comments (n 27) 939.
[34] Thym (n 3).
[35] Thym (n 3).
[36] Moravcsik (n 4).

influence does the individual citizen have?' Nor is it a psychological approach to whether citizens feel detached from the EU. These are perhaps legitimate concerns and not necessarily mitigated by the above analysis. Nonetheless such concerns do not only regard the EU but also national parliaments, this has been an attempt to conclude, overall, that the structural identity of the EU radically changes the question of democratic deficit.

APPORTIONMENT OF RESPONSIBILITY IN MEDICAL NEGLIGENCE

*PAUL WHITE**

Medical treatment should never break the chain of causation in order to allow the proper apportionment of responsibility and ensure legal certainty. The case of R v Jordan *(1956) 40 Cr App R 152 was wrongfully decided in so far as responsibility for death can never be justly absolved from the original assailant. Legal causation must take into account who is morally blameworthy. In conclusion, Parliament should abrogate medical negligence as a novus actus interveniens, as it would ensure legal clarity and proper apportionment of responsibility to the most culpable actor, with little de facto change to the law.*

The law in relation to medical negligence constituting a *novus actus interveniens* suffers from severe linguistic frailties. It displays little in the way of a clear doctrinal approach, allowing for policy considerations to lie beneath the vague terminology adopted by the judiciary. It is submitted that medical negligence should never break the chain of causation, in order to allow the proper apportionment of responsibility and ensure legal certainty. Furthermore, it is argued that *R v Jordan*[1] was wrongly decided and that the current law makes medical negligence breaking the chain of causation almost impossible. Therefore, Parliament should legislate to the effect that medical negligence may never break the chain of causation in criminal law.

The act of a third party breaks the chain of causation if it is free, voluntary and informed and it renders the original act no longer a substantial and operating cause of the outcome.[2] The courts have been reluctant to hold that medical negligence has broken the chain of causation in criminal law. However, in *Jordan* treatment which was described as 'palpably wrong'[3] was held to break the chain of causation. The two stab wounds, which had pierced the victim's intestines, had mainly healed at the time of death. Therefore, they could no longer be said to have caused the victim's death.

Later case law has marked a gradual retreat from Jordan. In *R v Smith*[4] the appellant stabbed the victim with a bayonet during a fight, piercing his lung and causing a haemorrhage. The victim subsequently died after receiving what was acknowledged as 'thoroughly bad'[5] medical treatment. However, the appeal was dismissed on the basis that the wound was still 'an operating cause and a substantial cause'[6] of death.

* Paul White, Newcastle University, Law LLB Stage Two.
[1] *R v Jordan* (1956) 40 Cr App R 152.
[2] *R v Smith* [1959] 2 QB 35, *per* Lord Parker C.J. 42.
[3] *Jordan* (n 1) 157.
[4] *R v Smith* [1959] 2 QB 35.
[5] *Smith* (n 4) 43.
[6] *Smith* (n 4) 42.

The circumstances in which medical negligence may break the chain of causation were further refined in *R v Cheshire.*[7] Cheshire shot the victim in the thigh and stomach, the wounds necessitated surgery from which he developed breathing difficulties. A tracheotomy was performed which led to scarring of the trachea, causing the victim to die. It was held that medical negligence would only break the chain of causation where it was 'so independent'[8] of the defendant's acts and 'so potent'[9] in causing death, that it rendered the defendant's contribution insignificant.[10]

Medical negligence should never break the chain of causation in criminal law, to allow otherwise would facilitate the possibility of improper attribution of moral responsibility. Responsibility for death can never be justly absolved from the original assailant. If the 'but for' test[11] is applied; but for the original attack, medical treatment would not have been necessary, thereby obviating any medical negligence. It is conceded that the 'but for' test is overly wide and inclusive. However, an original wounding is the only reason that treatment and presence in a hospital would be required. This holds no matter how gross the negligence or how much the original wound may have healed at the time of death. Responsibility must therefore be properly attributed to the defendant by never allowing medical negligence to break the chain of causation.

Tadros argues that 'causal enquiry is sensitive to moral factors.'[12] This is undoubtedly correct as fundamental moral principles underpin the criminal law. Legal causation must take account of who is morally blameworthy, indeed Williams describes legal causation as a test of 'moral reaction.'[13] The doctor will never be the most blameworthy. No matter how negligent they may have been they would still have been trying to help the victim recover from wounds inflicted by the defendant. Strawson suggests that to be responsible, a person must be an appropriate target for 'reactive attitudes' [14] in relation to their conduct. The role of doctors in society is to try to help, this is in stark contrast to the knife or gun wielding criminal who seeks to do harm. Consequently, the defendant will always be a greater target for negative reactive attitudes and, therefore, more morally responsible. It is not submitted that a doctor will not also be morally responsible and perhaps criminally liable for gross negligence manslaughter or damages in tort. However, the doctor will never be the most morally responsible and therefore medical negligence should never break the chain of causation.

In addition the law is too imprecise, leading to unacceptable degrees of uncertainty. The language used to define when a break in the chain of causation will occur is ambiguous and the case law is irreconcilable. While it is conceded that this affords the law greater

[7] *R v Cheshire* [1991] 1 WLR 844.
[8] *Cheshire* (n 7) *per* Beldam L.J 852.
[9] ibid.
[10] Ibid.
[11] *R v Dyson* [1908] 2 KB 454.
[12] Victor Tadros, *Criminal Responsibility* (OUP 2005) 156.
[13] Glanville Williams, *Textbook of Criminal Law* (2nd edn, Sweet & Maxwell 1983) 381.
[14] P. F. Strawson, *Freedom and Resentment and Other Essays* (Routledge 2008) 15.

flexibility to deal with novel cases this argument is more convincing in a private law context. However, it must be legal certainty which takes precedence in the criminal law, due to the restrictions on an individual's liberty which may be imposed. Abrogation of medical negligence as a *novus actus interveniens* would restore much needed certainty to the law.

In *Jordan* a distinction is drawn between normal and abnormal treatment. This is also advocated by Hart and Honoré.[15] However, it is vague and unhelpful. Norrie argues, '[t]he problem is that what is normal... is a matter of judgement and perspective.'[16]This is correct because what is perceived as 'normal' will vary enormously depending upon social background or geographical location. This ambiguity provides an intrinsically flawed base for the imposition of criminal sanctions.

Hart and Honoré also state that a *novus actus interveniens* must be a voluntary act,[17] this was considered to be 'broadly correct'[18] in *R v Pagett.*[19] However, this is yet another term which is inherently vague and promotes uncertainty in the law. Hart and Honoré refer to the American case of *State v Preslar,*[20] in which a woman died sleeping outside after leaving her house. They point out that her actions should be regarded as 'fully voluntary'[21] as she had slept outside without necessity. However, her choice takes on a much less voluntary appearance when it is considered that she was beaten by her husband. Her actions may more accurately be viewed as an act of necessity, induced by a desire for self-preservation. These differing conclusions highlight the difficulties with adequately defining what is meant by a term such as 'voluntary' and the consequent problems with legal certainty.

These linguistic frailties leave this area of causation underpinned by 'concepts which lack a valid theoretical grounding.'[22] The equivalence of the language leads to a wide variance in interpretation depending on whether focus is on individualism or a broader view of the social context. The emphasis on individualism and a narrow approach leaves causation analysis 'fundamentally flawed'[23] as it is left to policy to mark out the boundaries which the ambiguous language cannot. Public policy should not be veiled behind a curtain of vague terminology. It is undesirable for both legal certainty and it confers too wide a discretion upon the judiciary facilitating ex post facto rationalisation. Decisions may be reached based on policy and then justified by working backwards using vague terminology such as 'voluntary' or 'normal'. If medical negligence could not constitute a *novus actus interveniens* these deficiencies would be remedied.

---

[15] Hart and Honoré, *Causation in the Law* (2nd edn, OUP 1985) 340.
[16] Alan Norrie, *Crime, Causation and History: A Critical Introduction to Criminal Law* (2nd edn, CUP 2006) 138.
[17] Hart and Honoré (n 15) 352.
[18] (1983) 76 Cr App R 279, 289 *per* Goff L.J.
[19] (1983) 76 Cr App R 279.
[20] (1888) 50 Ark 545.
[21] Hart and Honoré (n 15) 327.
[22] Norrie (n 16) 141.
[23] Norrie (n 16) 140.

Furthermore, the case law cannot be reconciled. In both *Jordan* and *Cheshire* the wounds which the defendants had inflicted upon the victims were no longer life threatening nor were they the immediate cause of death. In *Jordan* the wounds were 'mainly healed at the time of death'[24] while in *Cheshire* the original bullet wounds 'no longer threatened the life of the deceased and the chances of recovery were good.'[25] The 'palpably wrong'[26] treatment in *Jordan* is analogous to the negligent lack of treatment in *Cheshire*. However, the cases were decided differently. In *Cheshire* it was held that medical negligence did not render the defendant's acts insignificant and consequently did not break the chain of causation. In *Jordan* the chain of causation was broken due to the 'not normal treatment.'[27] This incongruence in the law leads to little clarity as to when medical negligence will break the chain of causation. *Jordan* and *Cheshire* came to antithetical conclusions on similar facts. This presents problems with legal certainty, which would be alleviated if medical negligence could never break the chain of causation.

It is submitted that *Jordan* is an anomaly and should have been decided differently. A broader analysis illustrates that the only reason the victim received treatment was as a result of the original wound inflicted by the defendant.[28]It was therefore still a substantial cause of the death even if it had mainly healed. A broader analysis of *Jordan* is preferable to a narrow interpretation, as it gives a fuller and more accurate account of events and how they relate to each other. Moreover, applying the approach in *Cheshire*, it could not be said that the administration of antibiotics in *Jordan* was 'so independent' of the defendant's acts. It was intended to stop infection of the wound inflicted by the defendant, and therefore dependent upon the stabbing. If it is accepted that *Jordan* was wrongly decided, there are no reported cases of medical negligence breaking the chain of causation in English law.

More recent case law has moved away from *Jordan* and the current barriers to medical negligence constituting a *novus actus interveniens* are almost insuperable. In *Cheshire* it was held that medical negligence must be 'so independent' of the defendant's acts to break the chain of causation. It is submitted that all the medical treatment that a victim receives would be entirely dependent upon the injurious acts of the defendant. This assertion holds no matter how negligent the treatment may turn out to be. Wilson contends that a victim contracting food poisoning from hospital food would come within the 'so independent' requirement in *Cheshire*.[29] This claim is not correct. The wound would be both condition and cause of the victim being in hospital and eating hospital food. Patients do not check into hospitals and eat hospital food of their own

[24] *Jordan* (n 1) 156..
[25] *Cheshire* (n 7) 846.
[26] *Jordan* (n 1) 157.
[27] ibid.
[28] Norrie (n 16) 145.
[29] William Wilson, *Criminal Law* (2nd edn, Harlow: Longman 2003) 112-13.

volition. It is thus inaccurate to contend that this would be 'so independent' of the wounding as there is no other reason the victim would be eating hospital food unless they had been injured by the defendant. This highlights the difficulty in overcoming the 'so independent' criterion.

Moreover, a *novus actus interveniens* must be a voluntary act of a third party. However, it is doubtful that doctors act voluntarily in light of the decision in *Pagett*;[30] if this is so medical negligence could not break the chain of causation. In *Pagett*, police officers returning fire were held not to be acting voluntarily as they were acting under a duty, thus their actions could not break the chain of causation. The same should apply to doctors, who owe a duty of care to their patients. It would be somewhat contradictory to hold that the police act involuntarily while performing their duties while doctors act voluntarily in the performance of theirs.

It is submitted that Parliament should legislate to the effect that medical negligence can never break the chain of causation. The current linguistic barriers that negligence be 'so independent' and 'voluntary' mean that a legislative change of this kind would be more of a de jure change than a de facto one. Parliament would simply be elucidating what the law is in practice, as it is currently almost impossible for medical negligence to break the chain of causation. This would make patent the policy concerns underpinning the law, instead of allowing latent application of policy by the judiciary behind a veil of vague terminology. Thereby, ensuring clarity and more importantly the proper apportionment of responsibility to the most morally culpable.

Medical negligence breaking the chain of causation facilitates the injustice of absolving a defendant who is the most morally responsible for death. Furthermore, the vague terminology and contradictory case law fosters uncertainty and is an unsound basis for criminal sanctions. These factors highlight the need for Parliamentary intervention to abrogate medical negligence as a *novus actus interveniens*. Clarity and morally just apportionment of responsibility would be ensured, with little de facto change to the current law.

[30] *Pagett* (n 19).

ELECTRICITY PRODUCTION FROM INDUSTRIAL WASTE HEAT – IN THE BLIND SPOT OF EUROPE'S CLEAN ENERGY POLICIES?

*DIRK BÖEHLER**

In the face of climate change and diminishing reserves of fossil fuels, it is in the well understood interest of the world economy to reduce its carbon dependency. One important step towards this goal is to increase the efficiency of carbon-based processes by harnessing the waste heat energy that is generated through existing carbon-based energy generation methods. Historically, utilisation of waste heat energy for electricity production has been too expensive to be competitive; however with today's increasing energy prices and allocation rewards for efficient techniques in the European energy policies, the demand for this technology is high. This paper will focus on the complex regulatory framework that governs waste heat recovery in the form of electricity generation from waste heat in the ferro-alloy industry. The paper will analyse the position and treatment of waste heat electricity in the EU-ETS through an analysis of the provisions within.

1 INTRODUCTION

1.1 Electricity from Industrial Waste Heat

In the face of climate change and diminishing reserves of fossil fuels, it is in the well-understood interest of the world economy to reduce its carbon dependency. One important step towards this goal is to increase the efficiency of carbon-based processes. Most of these processes take advantage of the fact that coal, oil and natural gas contain large amounts of chemical energy. This energy is transformed into, *inter alia*, electric or thermal energy or chemical energy in another medium. The generation of the former can be the main purpose of the transformation (e.g. in residential heating or the supply of process heat for industrial processes), but more often, it is a mere by-product that cannot be avoided in the generation of one of the other forms of energy. Generally speaking, this excess heat is waste heat.

Using waste heat can contribute to the efficiency of the underlying processes significantly, as a greater energy yield requires less energy input for the same output. Utilisation for electricity production has in the past been too expensive to be competitive, and the infrastructure has been insufficient to enable comprehensive trade flows with heat as such. Recently, however, increasing energy prices and allocation rewards for efficient techniques in the European energy policies have created new incentives to invest in recuperation facilities.

As waste heat occurs in virtually every intentional energy transformation, the variety of possible applications is vast. This paper, however, will only deal with waste heat recovery in the form of electricity generation from waste heat in the ferro-alloy

industry, which is subject to a particularly complex regulatory framework and therefore in many respects exemplary for the difficulties occurring in the regulation of innovative technologies.

1.2 Legal Background

This paper will analyse the position and treatment of waste heat electricity in the European Union Emissions Trading System (EU-ETS). For that purpose, a great variety of legal documents will have to be taken into account. The trading scheme, which is informed by Article 17 Kyoto Protocol and the 5th EC Environmental Programme, is the largest emission trading scheme worldwide,[1] comprising 27 EU-Member States plus the EEA-States Iceland, Liechtenstein and Norway. The central document governing the scheme is Directive 2003/87/EC (ETS-Directive),[2] which has been amended by three further Directives and one Regulation;[3] all were preceded by the normal legislative procedure and are subject to the EEA Agreement, which provides, *inter alia*, for the promotion of equal conditions of competition.[4]

For reasons further explained in section 3, the most important ETS-provision for the following analysis is Article 10a, which provides for the establishment of rules for the free allocation of emission allowances to encourage the implementation of efficient technologies; these rules have been set by the European Commission in Decision 2011/278/EU (Benchmarking Decision).[5] The Decision was temporarily under challenge before the General Court of the European Union: The European steel industry association Eurofer argued, *inter alia*, that the Decision violates Article 10a ETS-Directive and infringes the principle of proportionality.[6] The Court, however, rejected the request due to Eurofer's lack of *locus standi*,[7] which is in line with previous case law.[8]

---

* Dirk Böehler, Newcastle University, LLM Environment and Sustainable Development.

[1] Josephine Van Zeben, 'Respective powers of the European Member State and Commission regarding emissions trading and allowance allocation' (2011) 12 ELR 216.

[2] Directive 2003/87/EC of the European Parliament and of the Council of 13 October 2003 establishing a scheme for greenhouse gas emission allowance trading within the Community and amending Council Directive 96/61/EC [2003] OJ L275/32.

[3] Directive 2004/101/EC of the European Parliament and of the Council of 27 October 2004 [2004] OJ L338/18; Directive 2008/101/EC of the European Parliament and of the Council of 19 November 2008 [2009] OJ L8/3; Regulation (EC) No 219/2009 of the European Parliament and of the Council of 11 March 2009 [2009] OJ L87/109; Directive 2009/29/EC of the European Parliament and of the Council of 23 April 2009 [2009] OJ L140/63.

[4] Agreement on the European Economic Area [1994] OJ L1/3, art 1(1).

[5] Commission Decision 2011/278/EU determining transitional Union-wide rules for harmonised free allocation of emission allowances pursuant to Article 10a of Directive 2003/87/EC of the European Parliament and of the Council [2011] OJ L130/1.

[6] Case T-381/11 *Eurofer v Commission* [2011] OJ C269/56.

[7] Case T-381/11 *Eurofer v Commission* [2012] (GCEU, 6 June 2012).

[8] Case C-263/02 P *Commission v Jégo-Quéré* [2004] (ECJ, 1 April 2004); Case 25-62 *Plaumann v Commission* [1963] (ECJ 15 July 1963); Case C-321/95 P *Stitching Greenpeace v Commission* [1998] (ECJ 2 April 1998), which, however, is not referred to in the decision.

Given the lengthy procedure preceding the preliminary ruling the association is now aiming for,[9] judicial review of the Benchmarking Decision cannot be expected anytime soon; instead, interpretation will have to be based on a number of Guidance Documents issued by the Directorate General on Climate Action (DG CLIMA) which provide interpretational assistance but are not legally binding.[10] The Guidance Documents relevant for this paper are number one, two, five and eight. The DG CLIMA clarified Guidance Document 8 in respect to waste gas recovery in a further Explanatory Note.

Further guidance is provided by the Reference Documents on Best Available Techniques (BREFs) and Impact Assessments (Staff Working Documents issued by the Commission) which inform the Decision.

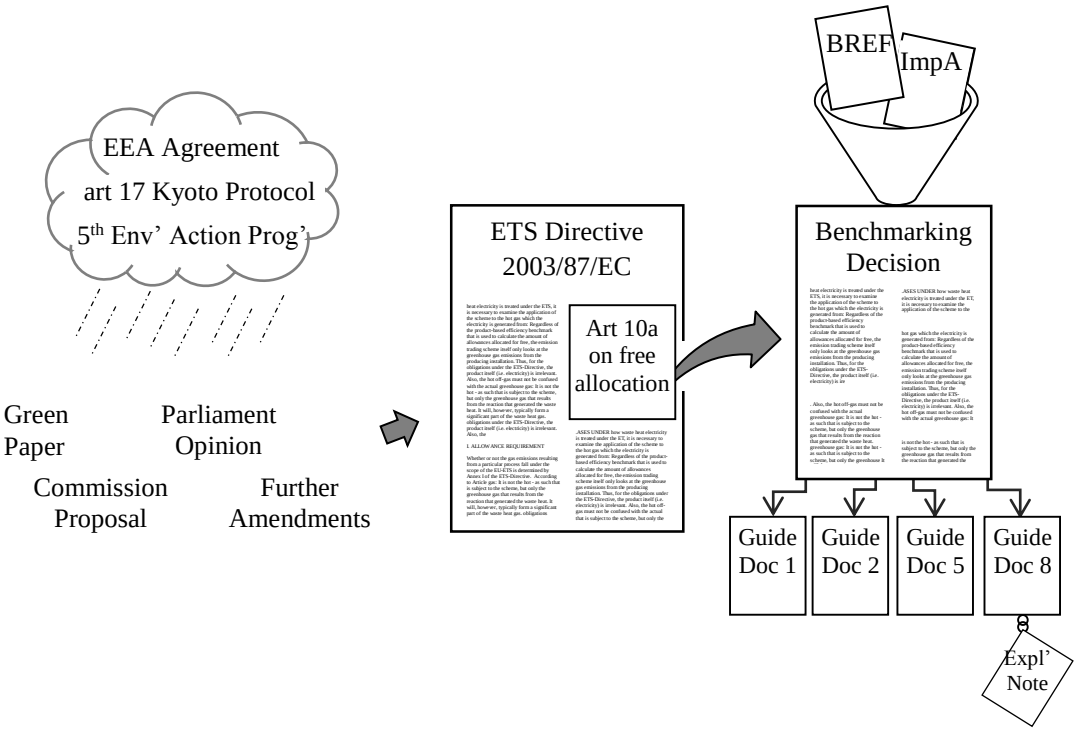


Diagram no. 1: Sources of Regulatory Framework for Industrial Waste Heat Electricity

This legal and regulatory framework provides the parameters for the following analysis of industrial waste heat electricity.

[9] Eurofer, 'Court case of European steelmakers on benchmarks under the ETS to be judged on national level' (*Eurofer*, 8 June 2012) <http://www.eurofer.eu/index.php/eng/News-Media/Press-Releases/Court-case-of-European-steelmakers-on-benchmarks-under-the-ETS-to-be-judged-on-national-level2> accessed 10 July 2012.

[10] DG-CLIMA, 'Guidance Document no1 on the harmonized free allocation methodology for the EU-ETS post 2012 – General Guidance to the allocation methodology' (Guidance Document 1) (2012) 3.

1.3 Sources and Methodology

Due to a lack of precedents, this paper will mostly rely on original research; as the Benchmarking Decision has been adopted only in 2011, there is no case law and virtually no scientific literature dealing with free allowance allocation available yet. Meanwhile, the emission trading schemes do not provide any codified rules in regards to energy recovery in the form of waste heat electricity. Against this background, the analysis will frequently refer to a real-life case to illustrate the current way the Commission interprets the schemes. These references do not aim to produce a case study, but help illustrate the abstract analysis with the condensate of the on-going discussion between the operator and the European Commission as an exemplary case. The installation under study is a Norwegian ferro-alloy plant operated by the medium-sized enterprise Finnfjord, which is currently being equipped with a waste heat electricity generator.[11]

1.4 Structure and Research Questions

The inquiry is structured as following: Sections two and three clarify the basic terms and mechanisms relevant for the further assessment: How is electricity from waste heat produced in the ferro-alloy industry? Why does the ETS provide for the allocation of free allowances, and what are the rules of this allocation? Based on these premises, the fourth section describes in detail how the ETS-Directive and the Benchmarking Decision are applied to processes generating industrial waste heat, and how to the electricity generated thereof. The fifth section evaluates this nexus: Is the way the schemes are interpreted compliant, reasonable and consistent? And does the scheme under its current interpretation by the Commission create a level playing field between waste heat electricity, heat delivery, renewable energy and conventional electricity—or is electricity from industrial waste heat in the blind eye of Europe' energy policies?

2 ELECTRICITY FROM INDUSTRIAL WASTE HEAT

To facilitate understanding of how waste heat electricity is approached under the European ETS, this section will describe the process of generating industrial waste heat and some technical aspects of the recovery of electric energy from therefrom.

2.1 Industrial Waste Heat

European legislation does not provide for a legal definition of waste heat. Therefore, it is hereby defined as 'heat that occurs as a by-product from mechanical, electrical or chemical processes that have a purpose other than the mere generation of heat.'

In the ferro-alloy production in so-called electric arc furnaces (EAF), large amounts of such heat is generated in the course of the reduction of metal oxides, which is the most important stage in the production process of ferro-alloys.[12] It takes place at high temperatures generated by electrodes using electric energy to heat up the ore in order

[11] Surveillance Authority EFTA, 'Decision of 9 February 2011 on the aid to Finnfjord AS for an energy recovery system' (Finnfjord Decision) (2011)(Dec No 39/11/COL) para 3.1.
[12] Commission, BREF Non Ferrous Metals (2001) 555.

to enable the chemical reactions in the furnace; therefore, the heat generated by the electrodes is not waste heat.[13] Depending on the product manufactured, different types of materials can be used as reducing agents to remove oxygen from the ore;[14] the production of ferrosilicon requires the use of carbon-bearing products, i.e. coal and coke.[15] During the process, the carbon-bearing reducing agents themselves release (thermal) energy in amounts that equal or even exceed the amount of electrical power used to trigger the reduction.[16] The result is hot off-gas that needs to be cooled or released to the surroundings in order to prevent damage to the production equipment.[17] A typical EAF for the production of ferro-silicon can be seen below:

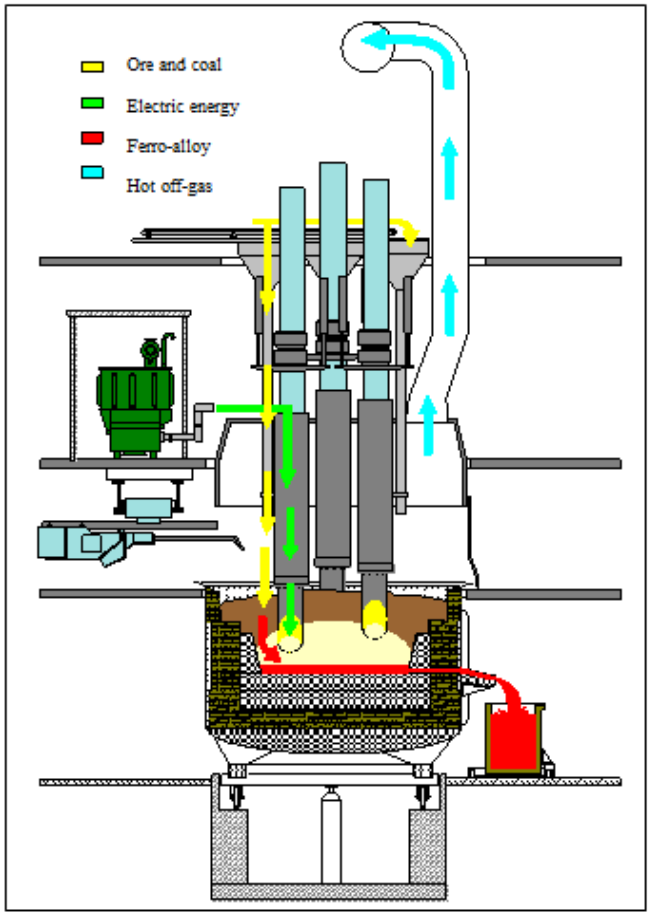


Diagram no. 2: Ferro-alloy production in the electric arc furnace.[18]

Although heat is required to run the reaction, the main purpose of adding the coal-bearing products to the process is to chemically reduce the metal oxides that are brought into the process as raw material. The energy released is therefore waste heat.

[13] DG-JRC, 'Draft Reference Document on Best Available Techniques for the Non-Ferrous Metals Industries' (2009) 635.
[14] ibid 635.
[15] Surveillance Authority EFTA, 'Decision of 9 February 2011 on the aid to Finnfjord AS for an energy recovery system' (Finnfjord Decision) (2011)(Dec No 39/11/COL) para 2.1.
[16] Finnfjord, *Licence Application for Thermal Power Station* (2010) 4.
[17] EFTA, Finnfjord Decision (2011) para 2.1.
[18] Commission, BREF Non Ferrous Metals (2001) 507 (The labels are edited).

As set out in the introduction, waste heat occurring in the ferro-alloy industry is the only form of waste heat that will be discussed within the scope of this paper as a source for electricity production. Hence, the term "industrial waste heat" will only refer to heat that is generated as described above.

2.2 Energy Recovery

Energy recovery from ferro-alloy smelting facilities is site-specific; whether it is feasible or not depends on factors such as the chemical consistence of the off-gas, the specific usage of heat energy in the production process and third-party heat demand.[19] Where the off-gas is rich in CO, the chemical energy content can be utilized by direct burning;[20] in other cases, energy can be recovered only through the utilization of the heat itself. As the burning of CO is independent from the temperature of the off-gas, only the latter form of energy recovery will be discussed as energy recovery from industrial waste heat.

Generally, the thermal energy content of industrial waste heat can be utilized in three different ways: (1) it can be recycled on site (heat recuperation), e.g. to dry or preheat the furnace charge; (2) it can be sold and transported via heat tubes to households or other industrial installations, so-called district heating and process heat; (3) it can be used for electricity production by a steam turbine on site.[21] The first two forms depend greatly on the process design of the specific site and patterns of third-party demand.

The third form of energy recovery (electricity production from industrial waste heat) produces a highly tradable product relatively independent from direct demand; therefore, it can be implemented where recuperation and district heating are not feasible. It uses boilers that convert the energy from the off-gases into overheated steam, which is then transformed into electric power through a turbine and a generator.[22] The electrical output can be sold, but it can also be used in the energy-intensive process of ferro-alloy production itself.[23]
It is this form of energy recovery that will be the subject matter of this paper.

[19] DG-JRC, 'Draft Reference Document on Best Available Techniques for the Non-Ferrous Metals Industries' (2009) 652; Commission, BREF Non Ferrous Metals (2001) 567.
[20] Commission, BREF Non Ferrous Metals (2001) 565.
[21] DG-JRC, 'Draft Reference Document on Best Available Techniques for the Non-Ferrous Metals Industries' (2009) 652; Commission, BREF Non Ferrous Metals (2001) xviii.
[22] EFTA, Finnfjord Decision (2011) para 4.1.2.
[23] ibid para 2.2.

3 THE EU REGIME GOVERNING THE ALLOCATION OF FREE ALLOWANCES

One of the obvious requirements of an emission trading scheme is the initial availability of tradable allowances.[24] The following section describes the methods used to allocate allowances to operators of industrial installations.

3.1 Auctioning and Free Allocation

Throughout the first and second trading period 2005-2012, Member States were allowed to auction up to ten per cent of allowances,[25] a right which they exercised only reluctantly.[26] The vast majority had to be – and was – allocated for free. But, from 2013 on, free allocation is technically supposed to be the exception: 'Member States shall auction all allowances which are not allocated free of charge'[27]. However, Article 10a(11) ETS-Directive provides that a share of 80% in 2013 (reducing by equal amounts to 30% in 2020) is to be allocated for free. The value determining the size of the share is called 'carbon leakage exposure factor' (CLEF),[28] as its purpose is to moderate the exposure of European operators to carbon leakage by introducing the scheme incrementally.[29]

Year	*2013*	*2014*	*2015*	*2016*	*2017*	*2018*	*2019*	*2020*
CLEF	0.800	0.729	0.657	0.586	0.514	*0.443*	*0.371*	*0.300*

The highlighted numbers suggest that, in terms of its share in actual distribution, free allocation will remain more common than auctioning for a few more years. This raises the question whether or not it is misleading to call free allocation an 'exception'.

The Commission estimates that as early as 2013, at least 50% of the total amount of allowances will in fact be auctioned.[30] This is possible because the CLEF does not relate to the total amount of allowances, but only to the 'quantity determined in accordance with the measures referred to in paragraph 1 (of Article 10a).'[31] These measures will be discussed in more detail later on; albeit, the quantity of allowances they determine is

[24] Stefan Weishaar, 'CO2 emission allowance allocation mechanisms, allocative efficiency and the environment: a static and dynamic perspective' (2009) 24 European Journal of Law & Economics 29, 36.
[25] ETS-Directive, art 10 in the version before its amendment by Directive 2009/29/EC.
[26] Anne Theo Seinen, 'State aid aspects of the EU Emission Trading Scheme: The second trading period' (2007) (3) Competition Policy Newsletter 100.
[27] ETS-Directive, art 10(1).
[28] DG-CLIMA, 'Guidance Document no1 on the harmonized free allocation methodology for the EU-ETS post 2012 – General Guidance to the allocation methodology' (Guidance Document 1) (2012) para 5.4.1.
[29] DG-CLIMA, 'Guidance Document no5 on the harmonized free allocation methodology for the EU-ETS post 2012 – Guidance on carbon leakage' (Guidance Document 5) (2011) para 3.1.
[30] Commission, 'Auctioning'.
[31] ETS-Directive, art 10a(11).

significantly lower than the total amount of emissions distributed. Therefore, a share of 80% of this quantity can amount for less than 50% of total allowances.[32]

However, the Directive provides that, in order to avoid carbon leakage, installations in sectors exposed to a significant risk thereof will be subject to a CLEF of 100% until 2020.[33]

3.2 Regulatory Implementation: Benchmarking and Fall-back Methods

3.2.1 Rules and Their Hierarchy

It has been stated above that in the third trading period of the emission trade, auctioning is to be the rule and free allocation the exception (Article 10(1) ETS-Directive). Hence, allowances are to be allocated for free only to a certain amount. This amount is the CLEF share (i.e. in 2013, 80%, see above) of the so-called 'preliminary annual number of emission allowances'.[34] The preliminary annual number of emission allowances of every installation is to be determined in accordance with the Benchmarking Decision, which has been adopted by the Commission pursuant to Article 10a(1)(1) of the Directive.[35] In principle, the Benchmark Decision provides guidelines for the exception to the rule of auctioning for all stationary installations under the ETS-Directive.[36] The power sector, however, 'given its ability to pass on the increased cost of carbon dioxide,'[37] is exempted from the exception (Article 10a(1)(3), last sentence); hence, there is no free allowance allocation for electricity production.

The main tool to determine the preliminary annual number of emission allowances for an installation is the benchmarking of product groups produced in the installations. This method is to be applied 'to the extent feasible,'[38] in cases where product benchmarks could not be established, fall-back approaches will apply.[39] In the following, both approaches will be examined.

3.2.2 Benchmarking

Benchmarks are performance standards for products used to determine the preliminary annual number of emission allowances.[40] The Commission's Benchmarking Decision establishes 'benchmarks for 53 industry product groups covering 75% of industrial

---

[32] Commission, 'Questions and Answers: Auctioning – general' (*European Commission*, 28 July 2011) <http://ec.europa.eu/clima/policies/ets/auctioning/third/faq_en.htm> accessed 5 April 2012.
[33] ETS-Directive, art 10a(12).
[34] DG-CLIMA, Guidance Document 1 (2012).
[35] Commission Decision 2011/278/EU determining transitional Union-wide rules for harmonised free allocation of emission allowances pursuant to Article 10a of Directive 2003/87/EC of the European Parliament and of the Council [2011] OJ L130/1.
[36] ibid, art 2.
[37] Directive 2009/29/EC of the European Parliament and of the Council of 23 April 2009 amending Directive 2003/87/EC so as to improve and extend the greenhouse gas emission allowance trading scheme of the Community [2009] OJ L140/63, rec (19).
[38] ETS-Directive, art 10a(1)(3).
[39] Benchmarking Decision (n 5), rec (12).
[40] Poncelet, 'The Emission Trading Scheme Directive: Analysis of Some Contentious Points', 6 European Energy and Environmental Law Review 246.

emissions under the EU ETS'[41]. The benchmark of a given product group is based on the average performance of the 10% most efficient installations in the respective sector.[42]

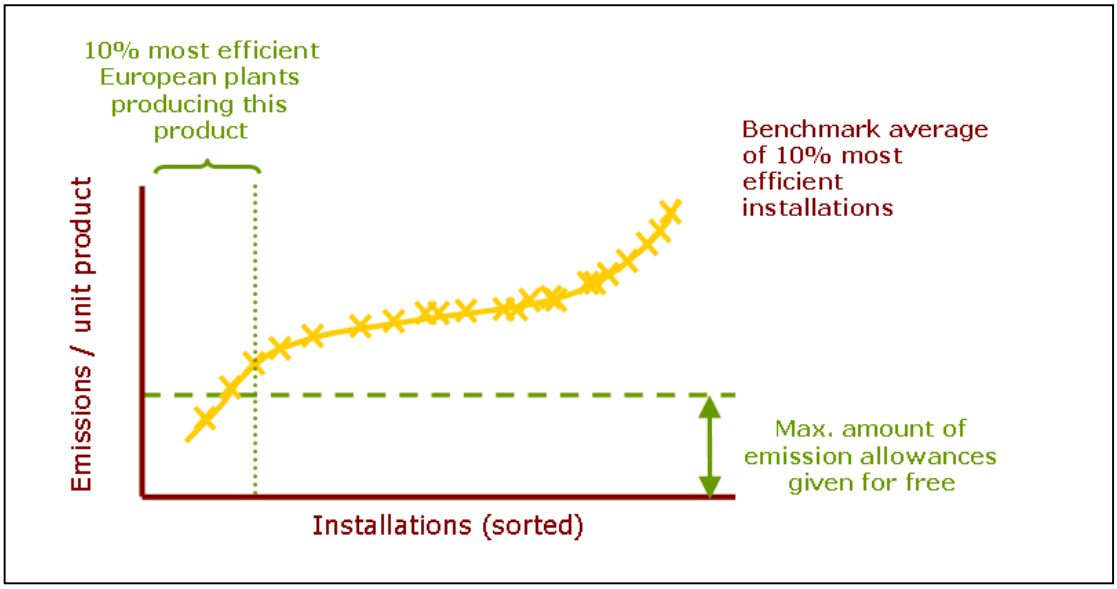


Diagram no. 3: Product Benchmarking[43]

'Installations that meet the benchmarks (and thus are among the most efficient installations in the EU) will in principle receive (the CLEF share of) all allowances they need.'[44] Hence, an operator O manufacturing *a product P* will receive for free 80% of the amount of metric tons of CO_2e emitted for the production of one P by the average 10% most efficient P-producers, times the number n of P that O produces. Note that, as the actual number n is still unknown by the time the allowances are allocated,[45] n is in fact not the number of P produced by O in the current year, but defined by historical activity levels as monitored during a baseline period.[46]

$$\textit{Number of allowances allocated free of charge for O} = \textit{CLEF} \times \underbrace{n \times \textit{tonnes of } CO_2e \textit{ emitted by the 10\% most efficient P-producers (on average) per P}}_{\textit{preliminary annual number of emission allowances}}$$

[41] ENDS, 'Commission finalises phase III EU ETS benchmark criteria' (2011) 432 ENDS Report 48.
[42] ETS-Directive, art 10a(2)(1).
[43] DG CLIMA, 'Guidance Document no1 on the harmonized free allocation methodology for the EU-ETS post 2012 – General Guidance to the allocation methodology' (2012) para 5.2.
[44] DG-CLIMA, 'Benchmarks for free allocation'.
[45] See ETS-Directive, art 11(2).
[46] Benchmarking Decision (n 5), art 9(2).

3.2.3 Fall-back Methods

In cases where time or data or the number of installations concerned was insufficient to establish benchmarks on a product basis,[47] the preliminary annual number of emission allowances is determined by a hierarchy of fall-back methods, namely (i) the heat benchmark, (ii) the fuel benchmark and (iii) process emissions.[48]

The heat benchmark applies in cases where no product benchmark has been defined to determine the preliminary annual number of emission allowances, but measurable heat is either consumed for a specific purpose (e.g. production or heating) in an ETS installation, or produced in an ETS installation and consumed by a non-ETS installation (to avoid double counting).[49] 'The benchmark here for 100% free allocation is the efficient production of heat using low-carbon natural gas'[50], which is set at 62.3 tonnes of CO_2 per terajoule of heat energy consumed.[51] Again, the preliminary amount of allowances is the product of the benchmark times the historical activity level of the respective installation *n*.[52] Under the heat benchmark approach, however, the benchmark does not refer to a number of products produced, but to the amount of measurable heat consumed during the baseline period, expressed as terajoule per year.[53]

$$\textit{Number of allowances allocated free of charge for O} = CLEF \times n \times \underbrace{\textit{62.3 tonnes of } CO_2 \textit{ / terajoule of heat energy consumed}}_{\textit{preliminary annual number of emission}}$$

The fuel benchmark applies in cases where a product benchmark is not available and heat produced by fuel combustion is consumed for a certain purpose (i.e. heating, cooling, mechanical energy or production) in amounts not measurable.[54] The benchmark is set at 56.1 tonnes CO_2 per energy content of the fuel consumed, expressed in terajoule.[55] The historical activity level of the respective installation *n* refers to the amount of fuel consumed during the baseline period.[56]

---

[47] ENDS Report Environment Daily, 'Commission finalises phase III EU ETS benchmark criteria' < http://www.endsreport.com/index.cfm?go=27015> (accesed 4 May 2012).
[48] Benchmarking Decision (n 5), rec (12).
[49] Benchmarking Decision (n 5) art 3(c); see also DG-CLIMA, 'Guidance Document no2 on the harmonized free allocation methodology for the EU-ETS post 2011 – Guidance on allocation methodologies' (Guidance Document 2) (2011) para 1.5.
[50] ENDS (n 47); Commission, 'Accompanying document to the Commission Decision on determining transitional Union-wide rules for harmonised free allocation pursuant to Article 10a of Directive 2003/87/EC' (Staff Working Document) SEC(2010) para 4.1.2.
[51] Benchmarking Decision (n 5), Annex I(3).
[52] Benchmarking Decision (n 5), art 10(2)(b)(i).
[53] Benchmarking Decision (n 5), art 9(3).
[54] Benchmarking Decision (n 5), art 3(d).
[55] Benchmarking Decision (n 5), Annex I(3).
[56] Benchmarking Decision (n 5), art 9(4).

$$\text{Number of allowances allocated free of charge for } O = CLEF \times n \times \underbrace{56.1 \text{ tonnes of } CO_2 \text{ / terajoule of fuel consumed}}_{\text{preliminary annual number of emission}}$$

The so-called process emissions benchmark applies in cases where emissions are neither covered by a product benchmark nor by the heat or fuel benchmark approach.[57] Here, the amount of allowances allocated free of charge is based on historical emissions.[58] However, the value is reduced to 97% '[i]n order to ensure that the free allocation of emission allowances for such emissions provides sufficient incentives for reductions in greenhouse gas emissions and to avoid any difference in treatment'[59].

$$\text{Number of allowances allocated free of charge for } O = CLEF \times 0.97 \times \underbrace{\text{Historical process emissions}}_{\text{preliminary annual number of emission allowances}}$$

3.3 Summary

Based on the general premise of a gradual reduction of the total number of allowances available, the portrayed scheme applies a finely tuned system of hierarchical methods that are linked to best practice as far as such a link was deemed feasible, and which refer to historical emission levels only as a fall-back approach. At the same time, it is a dubious masterpiece of European technocracy and bureaucracy which is difficult to apply to a specific case. Against this background, the next section will analyse how the ETS system is applied to waste heat electricity.

4 APPLICATION OF THE ETC SCHEME TO WASTE HEAT ELECTRICITY

4.1 Waste Heat Gases Under the ETS

The generation of waste heat electricity does not cause the emission of any additional greenhouse gases. Therefore, it is not directly subject to the emission trading scheme. What has to be examined instead is the application of the ETS to the hot gas the electricity is generated from: Regardless of the product-based efficiency benchmark that is used to calculate the amount of allowances allocated for free, the emission trading scheme itself only looks at the greenhouse gas emissions from the producing installation. In that context, it should be stressed that the hot off-gas (i.e. the waste heat

[57] DG-CLIMA, Guidance Document 2 (2011) para 1.5.
[58] Benchmarking Decision (n 5), art 10(2)(b)(iii).
[59] Benchmarking Decision (n 5), rec (12).

gas) as a whole is not identical with the actual greenhouse gas the operator must surrender allowances for. The latter will, however, typically form a significant part of the former.[60]

4.1.1 Allowance Requirement for Industrial Waste Heat Gases

Whether or not the greenhouse gas emissions resulting from a particular process fall under the scope of the EU-ETS is determined by the schedule in Annex I of the ETS-Directive.[61] According to Article 12(3), only operators of installations undertaking any activity set out in that list need to acquire (and later, surrender) allowances. Annex I also specifies the greenhouse gases emitted in a given processes which allowances have to be acquired for: The emission of other greenhouse gases (which are typically occurring in negligible quantities) is 'free'.[62]

The schedule lists, *inter alia*, the '[p]roduction or processing of non-ferrous metals, including production of alloys' and '[p]roduction or processing of ferrous metals (including ferro-alloys)'.[63] Ferro-alloys are alloys that contain iron; they are, however, non-ferrous (i.e. non-ferrous ferro-alloys) if the iron content is less than 50%.[64] The ferrosilicon produced by Finnfjord, for instance, is a non-ferrous ferro-alloy with an iron content of less than 25%.[65] Ferrous ferro-alloys have an iron share of more than 50%. Pure metals are non-ferrous if they do not contain iron at all;[66] otherwise, they are ferrous. As the lists covers both the production of ferrous and non-ferrous metals, all processes generating industrial waste heat within the meaning specified in section 2 fall within the scope of the ETS.

The heat energy that is released with the process off-gases stems from the reaction of carbon and oxygen to CO and CO_2, whereby the hot CO further oxidises to CO_2 in the presence of air.[67] As a result of these reactions, considerable amounts of carbon dioxide are generated and emitted.[68] Other greenhouse gases are released from the process as well;[69] their share, however, is limited. Therefore, the only greenhouse gas specified by the list in Annex I in respect to the activities identified above is carbon dioxide.

---

[60] DG-JRC, 'Draft Reference Document on Best Available Techniques for the Non-Ferrous Metals Industries' (2009) 632.
[61] Alfred Endres and Cornelia Ohl, 'Kyoto, Europe? An economic evaluation of the European Emission Trading Directive' (2009) 19 European Journal of Law & Economics 17, 18.
[62] See ETS-Directive (n 2), art 3(b).
[63] ibid, Annex I.
[64] Encyclopædia Britannica, 'Ferroalloy' (*Encyclopedia Britaanica*) <http://www.britannica.com/EBchecked/topic/205090/ferroalloy> accessed 5 May 2012.
[65] EFTA, Finnfjord Decision (n 11) para 3.4.
[66]Engineer's Handbook, 'Non-Ferrous Metals' (*EngineersHandbook*, 2006) http://www.engineershandbook.com/Materials/nonferrous.htm> accessed 4 May 2012.
[67] DG-CLIMA, 'Guidance Document no8 on the harmonized free allocation methodology for the EU-ETS post 2011 – Waste gases and process emissions sub-installation' (Guidance Document 8) (2011) 5.
[68] DG-CLIMA, 'Explanatory note on the definition of process emissions sub-installations' (Explanatory Note Process Emissions) (2011).
[69] T Lindstad and others, 'Greenhouse Gas Emissions From Ferroalloy Production' (INFACON XI 2007).

Article 12(3) obliges operators to surrender a number of allowances 'equal to the total emissions from that installation during the preceding calendar year'. It follows that allowances have to be surrendered for all CO_2 emitted in the process of generating industrial waste heat.

4.1.2 Free Allowance Allocation for Industrial Waste Heat Gases

4.1.2.1 Product Benchmark

It must now be determined to what extent these emissions are eligible for free allowance allocation. As described in the previous section, the Benchmarking Decision provides that, in principle, allowances should be allocated for free in accordance with a product benchmark. However, a survey informing the Benchmarking Decision concluded that:

> [f]or the (…) non-ferrous metals industry, the relatively small size of the sector in combination with the limited number of installations producing individual products resulted in the proposal not to cover these sectors via product benchmarks, but to apply the fall-back approaches.[70]

The Commission followed this proposal and refrained from developing a product benchmark for non-ferrous metals, including ferrosilicon. Consequently, these metals are not subject to any product benchmarks.

4.1.2.2 Heat Benchmark

In cases where no product benchmark applies, free allowances are allocated according to the heat benchmark. As an additional requirement, the installation needs to consume (a) measurable heat for (b) a specific purpose; alternatively, measurable heat can be exported to a non-ETS installation.[71] Measurable heat is defined as 'a net heat flow transported through identifiable pipelines or ducts using a heat transfer medium, such as, in particular, steam, hot air, water, oil, liquid metals and salts, for which a heat meter is or could be installed'[72]. The required conditions allow the heat energy to be treated in measurable quantities, making it analogous to a countable product.[73]

Although such a measurable heat carrier (a) does exist within the processes discussed in this paper, it does not meet the purpose criteria (b) of the heat benchmark as specified in Article 3(c) of the Benchmarking Decision; while some energy escapes through radiation and convection, or remains in the form of heat energy in the end product, or is converted into chemical energy in the end product components, heat energy is carried

[70] Ecofys, Fraunhofer Institute for Systems and Innovation Research and Öko-Institut, 'Methodology for the free allocation of emission allowances in the EU ETS post 2012' (2009) vii.
[71] Benchmarking Decision (n 5), art 3(c).
[72] Benchmarking Decision (n 5), art 3(e).
[73] Commission, 'Accompanying document to the Commission Decision on determining transitional Union-wide rules for harmonised free allocation pursuant to Article 10a of Directive 2003/87/EC' para 4.1.2.

in identifiable pipelines or ducts solely by the off-gases and cooling water.[74] The diagram below shows the energy flow in a 10 MW furnace for the production of silicon-metal.

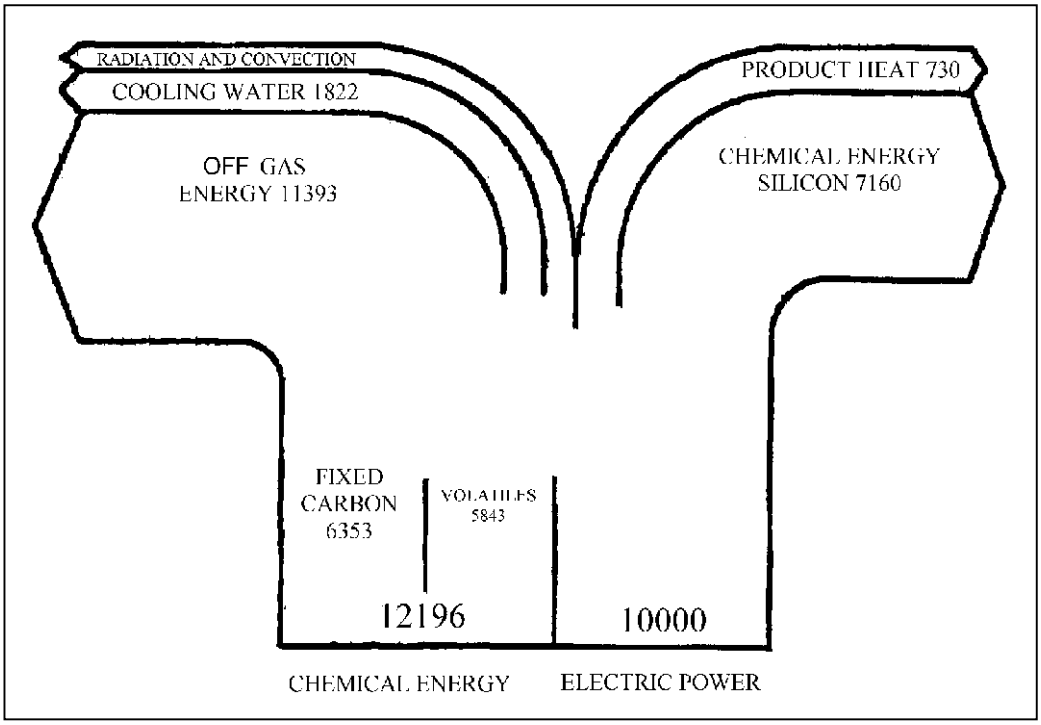


Diagram no. 4: Energy flow in Silicon-Metal production in Electric Arc Furnace.[75]

This measurable heat, however, is not 'consumed (…) for the production of products, for the production of mechanical energy (or) for heating or cooling'[76]; instead, it is merely removed from the furnace to prevent heat-related damage to the installation. In that context, the cooling water *transports* heat for the purpose of cooling, but the heat is not *consumed* for this purpose as, for example, in the utilization of the resulting steam pressure to power the gas compressor of a refrigerating system.

Hence, the installations discussed in this paper do not consume measurable heat for the purposes specified in the Benchmarking Decision; they are not subject to the heat benchmark.

4.1.2.3 Fuel benchmark

Where a product benchmark is not available and heat produced by fuel combustion is consumed for a specific purpose in amounts not measurable, the fuel benchmark applies.[77] 'Fuel' is not defined in the Benchmark Decision or any accompanying document, but one of the Impact Assessments informing the drafting of the latest

[74] Commission, BREF Non Ferrous Metals (2001) 546.
[75] Anders Schei, Johan Tuset and Halvard Tveit, *Production of High Silicon Alloys* (Tapir Forlag 1998).
[76] Benchmarking Decision (n 5), art 3(c).
[77] Benchmarking Decision (n 5), art 3(d).

amendment to the ETS-Directive clearly distinguishes between emissions 'released to the atmosphere as a consequence of combustion for the purpose of generating electricity and/or heat or as a consequence of another chemical process.'[78] Also, Guidance Document 8 issued by the DG CLIMA provides that '[f]uel used as reducing agent or for chemical syntheses should not be considered as fuel input into a fuel benchmark sub-installation.'[79]

Hence, coal used in the ferro-alloy production in an electric arc furnace does not constitute fuel within the meaning of the fuel benchmark provision; the fuel benchmark does not apply.

4.1.2.4 Process Emissions

Where neither the product benchmark nor any of the other fall-back approaches apply, free allowances are grandfathered according to the so-called process emissions approach: 'the preliminary annual number of emission allowances allocated free of charge for a given year shall correspond to the process-related historical activity level.'[80]

Subject to the requirement that they occur within an ETS installations, but outside the boundaries of a product benchmark, the emissions counting toward the historical activity level under this approach are listed in Article 3(h) and divided into three categories:

(a) non-CO_2 greenhouse gas emissions,
(b) CO_2 emissions from certain activities listed further below, and
(c) emissions from the combustion of incompletely oxidised carbon such as CO emitted by any of the following activities, if it is combusted to produce heat or electricity.[81]

As discussed previously regarding the activities generating industrial waste heat, the only greenhouse gas which allowances have to be surrendered for is carbon dioxide; therefore, type (a) emissions are irrelevant for the purposes of this paper. Type (c) emissions stem from the utilisation of the chemical energy content of the off-gases; this form of energy recovery does not fall within the scope of this inquiry, as it does not recover the waste heat, but the waste chemical energy. Also, the carbon dioxide in the hot off-gas does not stem from the utilization of the energy content of the off-gas, but from the process generating the off-gas itself. Therefore, the only emissions from installations generating industrial waste heat that count toward the historical activity

[78] Commission, 'Accompanying document to the Proposal for a Directive of the European Parliament and of the Council amending Directive 2003/87/EC so as to improve and extend the EU greenhouse gas emission allowance trading system' (Staff Working Document) SEC(2008) 52, 21.
[79] DG-CLIMA, Guidance Document 8 (2011) 16.
[80] Benchmarking Decision (n 5), art 10(2)(b)(iii).
[81] DG-CLIMA, Guidance Document 8 (2011) 5.

level are type (b) process emissions, i.e. CO_2 emissions from one of the following activities:

i. the chemical or electrolytic reduction of metal compounds,
ii. the removal of impurities from metals and metal compounds,
iii. the decomposition of carbonates,
iv. chemical syntheses where the carbon bearing material participates in the reaction, for a primary purpose other than the generation of heat,
v. the use of carbon containing additives or raw materials for a primary purpose other than the generation of heat,
vi. the chemical or electrolytic reduction of metalloid oxides or non-metal oxides such as silicon oxides and phosphates.[82]

Until now, the Court of Justice has not had any opportunity to comment on the interpretation and application of these categories. However, there is no apparent reason for assuming that they are meant to be mutually exclusive; consequently, a process can fall into more than one of the mentioned categories. The utilization of coal as reducing agent for the chemical reduction of iron oxide as described in section 2 lies within the scope of the last three categories. It follows that, in general, the CO_2 emitted from the processes generating industrial waste heat is considered a process emission eligible for free allocation.

However, not all of the carbon dioxide in the off-gas counts toward the historical activity level. This is because type (b) process emissions only comprise CO_2 emissions which occur 'as a result of any of the (listed) activities'[83]. According to the Guidance Documents issued by the Directorate-General Climate Action, the causal link between activity and emission has to be very close:

> Process emissions of type (b) only cover CO_2 as direct and immediate result of the production process or chemical reaction and as directly released to the atmosphere. CO_2 from the oxidation of CO is not covered by type (b) regardless if this oxidation takes place in the same or a separate technical unit. Example: A chemical reduction process leads to the production of a mix of CO and CO_2. At the presence of air, the CO is further oxidised to CO_2 and as result, 100% CO_2 is released to the atmosphere. The CO_2 from the oxidation of CO cannot be regarded as process emission type (b), since only the CO_2 as direct result of the activities can be considered as process emission of type (b) and as CO_2 from the oxidation of CO is covered by type (c) (in case of energy recovery).[84]

[82] Benchmarking Decision (n 5), art 3(h).
[83] Benchmarking Decision (n 5), art 3(h).
[84] DG-CLIMA, Guidance Document 8 (2011) 5 (parts omitted).

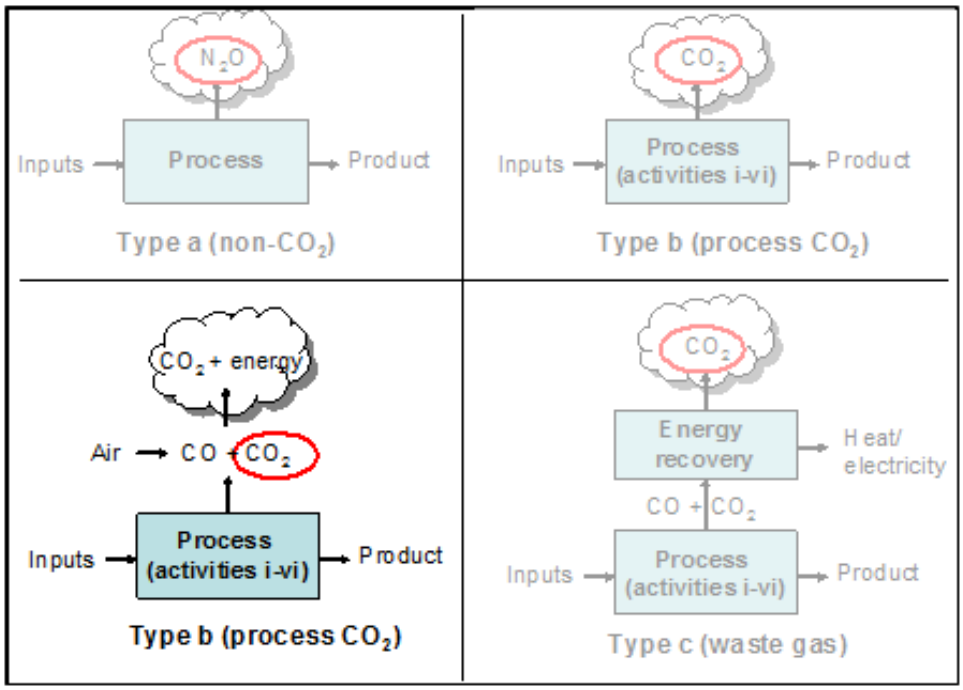


Diagram no. 5: Overview of process emissions installations (the emissions considered process emissions are marked by the red ellipses).[85]

It has been stated previously that in the processes generating industrial waste heat, CO occurs, but reacts further to CO_2. Hence, the processes are best described by the diagram on the bottom left. The only emissions that count toward the historical activity level (determining the amount of emissions eligible for free allocation) are the ones marked in red. As 'in practise no measurement data seems to be available'[86] on the share of the marked CO_2 in the total off-gas, 'a default value based on the assumption that 75% of the carbon content of the gas-mix is fully oxidised (CO_2) should be applied.'[87]

It follows that the historical process emission level of the relevant installations is 75% of their total CO_2 output. The preliminary annual number of emission allowances is 0.97 of that amount, or 72.75% of their total historical CO_2 output. As the production of non-ferrous metals has been deemed exposed to carbon leakage,[88] this preliminary number will not be modified by the CLEF until 2020.[89]

4.2 Waste Heat Electricity Under the ETS

Having figured out the amount of allowances allocated for free as regards the processes *generating* industrial waste heat, the crucial question is now whether this amount is further modified as a consequence of the *utilization* of this heat for the production of

[85] DG-CLIMA, Guidance Document 8 (2011) 7.
[86] DG-CLIMA, Explanatory Note Process Emissions (2011).
[87] DG-CLIMA, Guidance Document 2 (2011) 22.
[88] Commission Decision 2010/2/EU of 24 December 2009 determining, pursuant to Directive 2003/87/EC of the European Parliament and of the Council, a list of sectors and subsectors which are deemed to be exposed to a significant risk of carbon leakage [2010] OJ L1/10, Annex 1.1.
[89] ETS-Directive (n 2), art 10a(12).

electricity. For that purpose, both allowance requirement and free allocation eligibility need to be reviewed specifically in the light of the additional power production.

4.2.1 Allowance Requirement

Electricity production as such is not an activity listed in Annex I of the ETS-Directive. Therefore, producers only have to surrender allowances if they perform what is to be considered a case of – the listed – 'combustion of fuels'[90]. According to Article 3(t) of the Directive, '"combustion" means any oxidation of fuels, regardless of the way in which the heat, electrical or mechanical energy produced by this process is used'. As discussed previously, the carbon used for the reduction of ore does not constitute fuel; therefore, electricity production from waste heat does not impose an additional obligation on the operator. This makes sense, as the electricity generation does not increase the amount of emissions from the respective installation; also, energy recovery would otherwise be penalised rather than rewarded.

4.2.2 Free Allocation

4.2.2.1 Electricity Production in General

While electricity from industrial waste heat is not subject to an individual allowance surrender obligation, its production might influence the amount of free allowances allocated to the installation as a whole. This general assumption follows from the consideration that if two different products (i.e. alloys and electricity) are produced jointly, two different benchmarks for free allocation should apply cumulatively – after all, two separate installations would be allocated free allowances separately as well. This consideration is mirrored by Article 6 of the Benchmarking Decision, according to which each installation eligible for the free allocation of allowances should be divided into one or more sub-installations if different benchmarks are required.

However, emissions stemming from electricity production are generally not eligible for free allocation. Directive 2009/29/EC, which amends the ETS-Directive in preparation for the third trading period, provides that 'full auctioning should be the rule from 2013 onwards for the power sector, taking into account its ability to pass on the increased cost of CO_2'.[91] Consequently, the Benchmarking Decision does 'not cover the free allocation of emission allowances related to the production or consumption of electricity.'[92]

4.2.2.2 Emissions Used for the Production of Waste Heat Electricity

One exception is made to this rule. To provide incentives for energy efficient techniques, *inter alia* the efficient energy recovery of waste gases that cannot be

[90] ETS-Directive (n 2), Annex I.
[91] Directive 2009/29/EC of the European Parliament and of the Council of 23 April 2009 amending Directive 2003/87/EC so as to improve and extend the greenhouse gas emission allowance trading scheme of the Community [2009] OJ L140/63, rec 19.
[92] Benchmarking Decision (n 5), rec 31.

avoided in the industrial process,[93] installations combusting these waste gases for electricity or heat production receive allowances for free. The carbon dioxide occurring as a result of that combustion is, however, by definition a type (c) process emission; the greenhouse gases stemming from processes generating industrial waste heat, on the contrary, are type (b) process emissions. Hence, the utilization of that heat for the production of electricity does, in principle, not increase the amount of allowances allocated to the installation for free.

However, the steel industry was able to negotiate an interpretation that treats emissions used for waste heat electricity like emissions used for waste gas electricity, making the occurring CO_2 eligible for free allocation of type (c) process emissions. An Explanatory Note issued by the DG CLIMA states:

In this context, it is important to underline that all kinds of energy recovery would lead to the treatment of the respective amount of CO2 emissions as type (c) process emissions sub-installation. This includes the recovery of heat from the off-gas stream of the furnace.[94] It follows that the type (c) rules on free allocation for waste gas electricity apply analogously to waste *heat* electricity, too, although the emissions used for the latter were originally regarded type (b) emissions.

However, not all emissions occurring in the combustion of waste gas are eligible for free allocation, as the ETS-Directive provides only for a limited exception from the rule that '[n]o free allocation shall be made in respect of any electricity production, except for (…) electricity produced from waste gases.'[95] Therefore, emissions are allocated only as far as they occur from the combustion of waste gas instead of from the combustion of (more efficient) natural gas; emissions stemming from the hypothetical combustion of an energy-equivalent amount of natural gas have to be subtracted. Guidance Document 8 provides for a formula to calculate the historical activity level (HAL):

$$\begin{array}{c}\textit{Historical}\\ \textit{emissions}\\ \textit{(in t } CO_2\textit{e)}\end{array} = \begin{array}{c}\textit{Amount of}\\ \textit{waste gas}\\ \textit{(in tonnes)}\end{array} \times \begin{array}{c}\textit{Net calorific value}\\ \textit{of the waste gas}\\ \textit{(in TJ/t)}\end{array} \times \left[\begin{array}{c}\textit{Emission factor}\\ \textit{of the waste gas}\\ \textit{(in t } CO_2\textit{/TJ)}\end{array} - \begin{array}{c}\textit{Emission factor}\\ \textit{of natural gas}\\ \textit{(in t } CO_2\textit{/TJ)}\end{array} \times \textit{Correction } \eta\right]$$

The subtracted reference emissions are modified by correction factor ŋ to prevent market distortion at the expense of waste gas/heat electricity producers: It 'accounts for the difference in efficiencies between the use of waste gas and the use of the reference fuel natural gas'[96], as 'most waste gases have a higher emission intensity and can

[93] ETS-Directive (n 2), art 10a(1)(3).
[94] DG-CLIMA, 'Explanatory note on the definition of process emissions sub-installations' (Explanatory Note Process Emissions) (2011).
[95] ETS-Directive (n 2), art 10a(1)(3).
[96] DG-CLIMA, Guidance Document 8 (2011) 13.

therefore be used less efficient compared to other fuels.'[97] The default value of this factor is 0.667 but can be specified further if sufficient data is available.

In case of waste heat electricity, the numbers inserted into the formula are partly hypothetical, as no chemical energy is actually being used; instead, the default mix of 75% CO_2 and 25% CO mentioned above in respect to type (b) process emissions applies again.[98] It should be noted that the 75% CO_2 will not be allocated as type (b) process emissions any longer, as '[t]he content of CO_2 in the waste gas is treated as part of the waste gas stream'[99] and allocated as type (c) emission accordingly.

The following calculation will determine the historical emissions for 100 tonnes of fully oxidised (i.e. pure CO_2) off-gas. As 25% of this off-gas are considered CO, which is considerably lighter than CO_2,[100] the amount of waste gas is, for the purpose of the calculation, 90.91 t.[101] The net calorific value is equal to the energy content of CO,[102] but reduced in correspondence to the share of CO in the total off gas.[103] The emission factor of waste gas is in principle the weight ratio of CO and CO_2 divided by the energy content of CO. However, for every tonne of CO "combusted", a large amount of primary CO_2 is emitted, too,[104] increasing the emission factor of the waste gas to 621.71 t CO_2/TJ.[105] The emission factor of natural gas is set at 56.1 t CO_2/TJ.[106]

$$\textit{Historical emissions (in t } CO_2e) = 90.91\ t \times 0.00177\ TJ/t \times \left[621.71\ t\ CO_2/TJ - 56.1\ t\ CO_2/TJ \times 0.667\right] = 94.05\ t$$

As almost all variables can be calculated in dependency to the total amount of off-gases (and hence be eliminated), the complete formula to calculate the historical activity level of y tonnes of off-gas is $\frac{10}{11}$ y times 1.0334. It follows that, after the elimination of rounding errors, 93.95 tonnes of 100 tonnes of off-gas are considered type (c) process emissions, of which 97% (91.14 tonnes) are allocated for free.[107] Against the 75 tonnes (97%: 72.75 tonnes) that would have been recognized as type (b) process emissions and allocated for free without any heat recovery, this is a significant increase.

---

[97] DG-CLIMA, Guidance Document 8 (2011) 5.
[98] See DG-CLIMA, Guidance Document 2 (2011) 22.
[99] DG-CLIMA, Guidance Document 8 (2011) 13 f.
[100] C: 12 g/mole; O: 16 g/mole. Hence, CO: 28 g/mole; CO_2: 44 g/mole.
[101] 75 t + 25 x $\frac{7}{11}$ t = 75 t + 15.91 t.
[102] 10.1068 MJ/kg, or 0.01011 TJ/t; 'Fuel Gases – Engineering Toolbox, 'Heating Values' (*The Engineering Toolbox*) <http://www.engineeringtoolbox.com/heating-values-fuel-gases-d_823.html> accessed 10 July 2012.
[103] 10.1068 MJ/kg x $\frac{15.91}{90.91}$ = 1.77 MJ/kg (or 0.00177 TJ/t).
[104] $\frac{75}{15.91}$ tonnes.
[105] $\left(\frac{11}{7} + \frac{75}{15.91}\right) * 0.01011\ \mathrm{TJ/t^{-1}}$.
[106] Benchmarking Decision (n 5), Annex I(3).
[107] Benchmarking Decision (n 5), art 10(b)(iii).

4.2.3 Incidental Reduction of Allowance Payment

On account of its ability to pass on increased costs of CO_2 emissions, the power sector is generally not eligible for free allowance allocation. This means that, as far as they are unable to pass on increased costs as well, power consumers must pay for the allowances the electricity producers have to surrender. In fact,

> *[p]ower prices in EU countries have increased significantly since the EU emissions trading scheme (ETS) became effective (…) these increases in power prices may - at least in part - be due to this scheme, in particular due the pass-through of the costs of EU allowances (EUAs) to cover the CO_2 emissions.*[108]

Consequently, operators can reduce the internalised costs of allowances required for power generation by reducing the amount of grid electricity they purchase. That way, the generation of waste heat electricity reduces the amount of allowances operators have to pay for through the energy price.[109]

To quantify this reduction, the waste heat electricity production has to be set into context with the CO_2 emissions stemming from the grid power production. The mean EU-27 energy mix entails carbon emissions of 374.6g CO_2/kWh.[110] But since it is not the production of electricity as such but only the combustion of fuels which allowances have to be surrendered for, emissions from renewable energies (mostly construction) and nuclear power (uranium enrichment) have to be subtracted, resulting in about 350g CO_2/kWh.[111] However, the industry claims that, due to market mechanisms not further explained in this paper, this factor is to be set at around 750g CO_2/kWh.[112] Although the motive was clearly to project higher electricity costs and, that way, increase prospects of state aid, the Commission in principle adopted the argument;[113] consequently, the same transfer factor should be used for the reduction of indirectly paid allowances. Thus, for every 1.33 MWh of electric energy recovered from waste heat and used on site, one less allowance has to be paid for through the price for grid electricity.[114]

---

[108] JPM Sijm, SJ Hers, W Lise and BJHW Wetzelaer, *The impact of the EU ETS on electricity prices* (Energy Research Centre of the Netherlands 2008) 11.

[109] See also DG-CLIMA, 'Explanatory note on the definition of process emissions sub-installations' (Explanatory Note Process Emissions) (2011).

[110] Union of the Electricity Industry, *Power Statistics* (2010 edition, 2010) 15.

[111] Data based on best available technique; DG-JRC, 'Greenhouse Gas Emissions From Fossil Fuel Fired Power Generation Systems' (2001) 11; European Wind Energy Association, 'EU Energy Policy to 2050' (2011).

[112] Alliance of Energy Intensive Industries, 'Realistic evaluation of the indirect cost effects in the EU Emissions Trading Scheme for the analysis of the risk of carbon leakage and for financial compensation' (2009) 3.

[113] Commission, 'Impact Assessment Report: Guidelines on certain State aid measures in the context of Greenhouse Gas Emission Allowance Trading Scheme' (Staff Working Document) SWD(2012) 130 final Part 4, para 4.5.2.

[114] $1.33\ \mathrm{MWh/t\ CO2} = \frac{1}{0.750\ \mathrm{t\ CO2/MWh}}$.

It is important to understand, however, that this value is highly theoretical and varies greatly from case to case, as it depends on the location of the industrial installation and the energy mix of the respective power supplier. In Norway (which participates in the ETS but does not count toward the stated EU-27 energy mix), power generation is 98.5% hydro-based 'and as such, practically CO_2 free',[115] while 92% of Polish electricity is generated from coal.[116] Consequently, the Commission is calculating with a range of transfer factors between 400g and 900g CO_2/kWh.[117] It follows that, even though the European electricity grids in the central western area and the Nordic area are transnational,[118] the reduction of allowances indirectly paid for can still be expected to be much greater in Poland than in Norway.

4.3 Summary

The application of the EU-ETS and, specifically, of the rules governing the free allocation of allowances to installations generating industrial waste heat and electricity derived therefrom leads to the following result: The emissions occurring in processes generating industrial waste heat are, in principle, type (b) process emissions within the meaning of Article 3(h) Benchmarking Decision. This means that 72.75% are eligible for free allocation.

The additional production of electricity from that heat does not add to the amount of allowances that have to be surrendered; instead, it allows the emissions to be interpreted as type (c) process emissions with a 25% share of incompletely oxidised carbon, resulting in 91.14% of emissions covered by free allowances. Meanwhile, on-site consumption of waste heat electricity can replace conventional grid power and, that way, reduce payments for allowances passed through alongside the energy price by the power suppliers.

5 CRITICAL REVIEW OF THE SCHEME AND ITS APPLICATION

The analysis in section 4 raises questions about the validity of some provisions in the ETS-Directive and the Benchmarking Decision as well as about the authority of the understanding of these provisions as communicated in the Guidance Documents issued by the DG CLIMA.

The first question concerns the fact that ferro-alloy producers have no prospect of receiving 100% of the preliminary annual amount of allowances allocated for free. This might be a wrongful implementation of the ETS-Directive, which provides in Article

[115] Energy Norway, Federation of Norwegian Industries and Industri Energi, 'Carbon Price Transfer in Norway' (2011) 3.
[116] EFTA, Finnfjord Decision (2011) para 3.3.1; Commission, 'Poland – Energy Mix Fact Sheet' (2007) 2.
[117] Commission, 'Impact Assessment Report: Guidelines on certain State aid measures in the context of Greenhouse Gas Emission Allowance Trading Scheme' (Staff Working Document) SWD(2012) 130 final Part 4, para 4.5.2.
[118] Ibid, para 4.5.2.

10a that, in principle, the 10% most efficient producers in a certain sector should preliminarily receive all allowances they need. This section reviews the regulatory grounds of this allocation deficit and examines the individual factors causing the shortfall.

The allocation deficit is also the starting point of the next inquiry: The ETS-Directive provides that one of the purposes of free allocation is to avoid undue distortion of competition. Therefore, it needs to be evaluated if the rules on free allocation create a level playing field between waste heat electricity, heat delivery, electricity from renewable energies and conventional grid power.

5.1 Allocation Deficit

5.1.1 Regulatory Derivation

Operators of installations generating waste heat electricity receive free allowances according to the allocation method for type (c) process emissions. This method has to be understood in the context of Article 10a(1)(3) ETS-Directive, which provides that '[n]o free allocation shall be made in respect of any electricity production, except for (…) electricity produced from waste gases.'[119] In its initial proposal, the Commission did not envisage such an exception;[120] it was only inserted after various Members of Parliament pressed for an amendment of the proposed draft.[121] It is unclear why waste heat electricity remained unmentioned even though the MPs were obviously informed by the respective stakeholders. They might have considered it a special form of waste gas electricity; the DG CLIMA, however, adopted a very narrow interpretation of what is to be considered waste gas, including only

Gases which emerge from incomplete combustion or other chemical reaction in an EU-ETS installation and which comply with all of the following criteria:

1. Waste gases are not emitted without further combustion due to a significant content of incompletely oxidised carbon
2. The calorific value of waste gases is high enough for the waste gas to burn without auxiliary fuel input, or to contribute significantly to the total energy input when mixed with fuels of higher calorific value
3. The waste gas is produced as by-product of a production process.[122]

[119] ETS-Directive (n 2), art 10a(1)(3).
[120] Commission, 'Proposal for a Directive of the European Parliament and of the Council amending Directive 2003/87/Ec so as to improve and extend the greenhouse gas emission allowance trading system of the Community' COM(2008) 16 final, art 1(8).
[121] European Parliament, 'Report on the proposal for a directive of the European Parliament and of the Council amending Directive 2003/87/EC so as to improve and extend the greenhouse gas emission allowance trading system of the Community' (Session Document) A6-0406/2008; European Parliament, 'Amendments by Parliament to the Commission proposal for a Directive of the European parliament and of the Council amending Directive 2003/87/EC so as to improve and extend the greenhouse gas emission allowance trading system of the Community' A6-0406/180.
[122] DG-CLIMA, Guidance Document 8 (2011) 4 (numbering added).

Again, it could be asked why the definition was phrased in a way that excludes hot off-gases without chemical energy content; at the time, waste heat recovery was known and described in the BREF Document in great detail.[123] Nevertheless, the restrictive interpretation is in line with the narrow wording of the Directive.

However, the Commission did recognise that in the absence of chemical energy contents in the off-gas stream, the recovery of thermal energy does constitute the most efficient process design. As Article 10a(2) ETS-Directive provides that, in principle, the benchmark for free allocation should be set at 'the average performance of the 10% most efficient installations in a sector',[124] installations generating waste heat electricity should receive all allowances they need for free: 'Assuming product benchmarks for the production of certain ferro-alloys had been developed, the values of such benchmarks would have been determined by the most efficient installations, most likely applying energy recovery technologies.'[125] The extremely narrow interpretation of type (b) process emissions, however, limits the amount of emissions eligible for free allocation to 75% of the total off-gases. To bypass this inherent limitation, installations causing type (b) process emissions are, for the purpose of establishing the historical activity level eligible for free allocation, considered type (c) process emissions sub-installations if they recover (any) energy from the off-stream.[126] Given that the corresponding calculation method is designed for chemical energy recovery, the result is an entirely hypothetical activity level constructed from various default values and forced into shape through a questionable interpretation of the relevant provisions:

5.1.2 Ratio of Fully and Incompletely Oxidised Carbon

The hypothetical chemical energy content of the off-gas stream is calculated based on the assumption that the off-gas consists of 75% completely and 25% incompletely oxidised carbon.[127] In fact, however, 'no measurement data seems to be available'[128] that could confirm this value; arguably, it is a compromise between the position of the Commission, which aimed at a 50/50 ratio first,[129] and the industry, which claims that all carbon content is fully oxidised in the furnace already.

The argument first became relevant when the DG CLIMA decided that carbon dioxide is eligible for free allocation of type (b) process emissions only as far as it stems from the industrial process itself, but not as far as it stems from the oxidation of carbon monoxide; against this background, the industry obviously had an interest to press for a 'genuine' CO_2-share as high as possible.

[123] Commission, BREF Non Ferrous Metals (2001) 566.
[124] ETS-Directive (n 2), art 10a(2).
[125] DG-CLIMA, Explanatory Note Process Emissions (2011).
[126] DG-CLIMA, Explanatory Note Process Emissions (2011).
[127] DG-CLIMA, Guidance Document 2 (2011) 22
[128] DG-CLIMA, Explanatory Note Process Emissions (2011)
[129] DG-CLIMA, Explanatory Note Process Emissions (2011)

After it was agreed that occurring emissions qualify as type (c) process emissions under waste *heat* recovery, too, the ratio became relevant in a different context: One of the factors of the formula for calculating the historical activity level of type (c) process emissions is the emission factor of the waste gas, which is expressed in CO_2/TJ. With a larger share of carbon dioxide in the off-gas, the nominator increases, while its denominator shrinks alongside the reduced calorific value of the off-stream.[130]

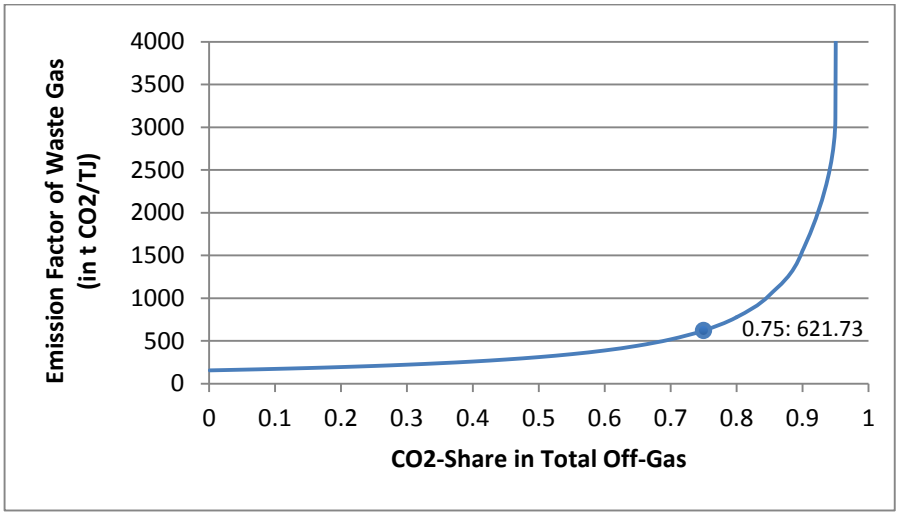


Diagram no. 6: Emission Factor versus CO_2 Share

The diagram above sets the emission factor into context with the share of fully oxidised carbon. The full formula modifies the equation in favour of a linear relation, but a higher share of carbon monoxide still results in a lower level of free allocation. The chart below demonstrates that with increasing chemical energy content and, thus, increasing potential for electricity production, installations generating waste gas experience a linear drop in allowance allocation (blue line); in case of a pure CO off-gas, only 74% of the resulting CO_2 would be allocated as type (c) process emissions.[131] Producers of waste heat electricity, meanwhile, receive a relatively high level of allowances regardless of the thermal energy content (and, thus, potential for electricity production) due to the fixed hypothetical CO/CO_2 mix (red line).

[130] The resulting formula for the emission factor of an off-gas with a CO-share x is $155.43 * \frac{1}{x}$ tCO_2/TJ

[131] $0.97 * \frac{7}{11}\ \mathrm{t} * 0.01011\ \mathrm{TJ/t} * \left(\frac{\frac{11}{7}}{0.01011}\ \mathrm{t\,CO2/TJ} - 56.1\ \mathrm{t\,CO2/TJ} * 0.667\right) = 0.74\ \mathrm{t\,CO2}.$

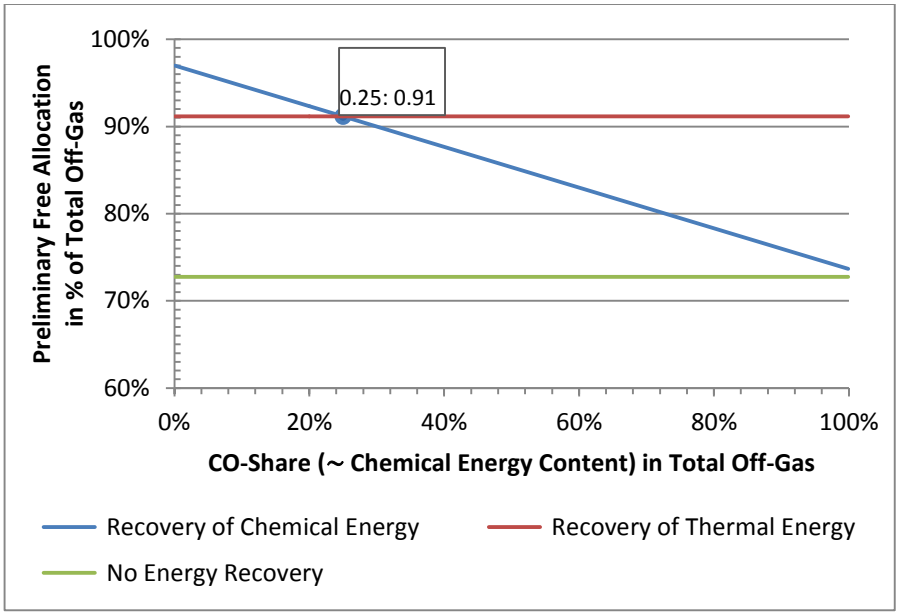


Diagram no. 7: Effect of Increasing Chemical Energy Content

5.1.3 Deduction of Hypothetical Natural Gas Emissions

However, the diagram above also illustrates that producers of waste heat electricity cannot receive all allowances they need; allocation only amounts to about 91%. This is the result of a conservative approach to the implementation of the waste-gas-exception from the rule of not allocating allowances to electricity production:

Given its initial reluctance to allocate free allowances to certain privileged forms of electricity production, it comes as no surprise that the Commission decided to limit free allocation to that part of the type (c) process emissions that occurs because the electricity is generated from waste gas *in lieu of* natural gas: 'Only emissions which are additional to the emissions that would occur if natural gas was used are taken into account.'[132] It follows that a hypothetical amount of emissions calculated on the basis of an energy-equivalent amount of combusted natural gas is subtracted from the historical activity level. The amount of natural gas needed to equal the energy content of the off-gas depends, again, on the share of incompletely oxidised carbon, which in case of waste heat recovery is set at 25%. The argument is that, as far as the chemical or thermal energy from the off-stream merely replaces the energy input of conventional electricity production, waste gas/heat electricity and conventional electricity should be treated equally. As electricity production is, in principle, not eligible for free allocation, equal treatment means "no allocation", which is implemented through the emission deduction.

It is not without doubt, though, if this method complies with the exception the Directive made for waste gas electricity: 'One could argue (…) that the text in Article 10a(1)(3) of the amended Directive (…) implies that for electricity producers using the waste gas,

[132] DG-CLIMA, Guidance Document 8 (2011) 5 (emphasis maintained).

(…) the total amount should be allocated.'[133] In that case, the chosen allocation method would infringe the provision. However, excluding electricity from free allocation was one of the leitmotifs of the 2009 amendment;[134] in that context, it is in line with accepted methods of legal interpretation to apply a restrictive approach towards any exception from that rule, and hence to subtract the hypothetical emissions from an equivalent amount of reference fuel.

The result, however, is that operators of installations generating industrial waste heat will never obtain full allocation: They can either refrain from energy recovery and receive allowances for type (b) process emissions amounting to 72.75% of the actual emissions only (green line); or they can recover energy and become eligible for type (c) process emission allowances, but then have to accept the reduction of their eligible activity levels to 91.14%.

5.1.4 Correction Factor ŋ

Treating conventional electricity and the equivalent part of waste gas / heat electricity equally has another paradoxical consequence: The more efficient energy recovery from the off-stream is, the lower is the historical activity calculated according to the formula. The chemical energy of natural gas can be transformed into electricity much more efficiently than the chemical or thermal energy content of off-gases. Hence, equivalent energy contents do not correspond to equivalent amounts of electricity; therefore, equivalent amounts of conventional electricity and waste gas/heat electricity should not be denied equal amounts of free allocation. As a consequence, the DG CLIMA introduced correction factor ŋ to reduce the amount of subtracted reference emissions; the factor 'accounts for the difference in efficiencies between the use of waste gas and the use of the reference fuel natural gas'.[135] The default value of this factor was negotiated to be 0.667, which means that waste gas/heat electricity is considered to be generated one third less efficiently than electricity from natural gas. If, however, 'the uses of waste gas and efficiencies related to these uses are known',[136] the factor can be adjusted.

It follows that energy recovery with a comparative efficiency deficit of less than one third would have to be assigned a higher correction factor, which would increase the amount of reference emissions and reduce the historical activity level eligible for free allocation. The logic applied is that with increased efficiency, the facility is more similar to a regular power producer, and, thus, fewer emissions are occurring due to the fact that waste gas instead of natural gas is burnt. In practice, however, it is unlikely

[133] Ecofys, Fraunhofer Institute for Systems and Innovation Research and Öko-Institut, 'Methodology for the free allocation of emission allowances in the EU ETS post 2012' (2009) 62.
[134] Directive 2009/29/EC of the European Parliament and of the Council of 23 April 2009 amending Directive 2003/87/EC so as to improve and extend the greenhouse gas emission allowance trading scheme of the Community [2009] OJ L140/63, rec (19).
[135] DG-CLIMA, Guidance Document 8 (2011) 13.
[136] DG-CLIMA, Guidance Document 8 (2011) 13.

that operators will actually produce data supporting such an unfavourable adjustment of the default value.

The diagram below compares allocation levels for recovery of thermal and chemical energy content based on the default correction value to allocation levels without correction, i.e. based on maximum efficiency of the electricity production.

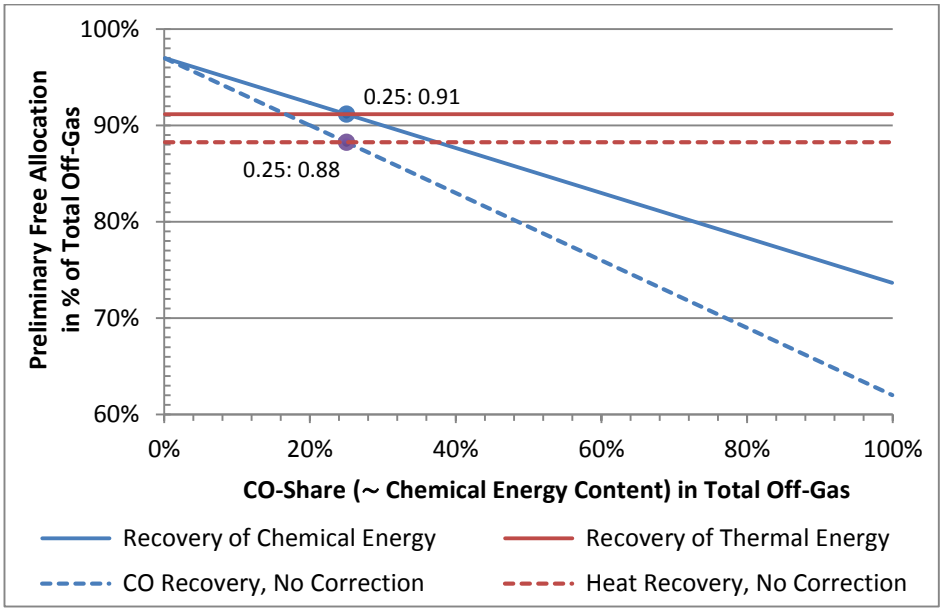


Diagram no. 8: Effect of the Correction Value on Allocation Levels

The result is particularly interesting in context with Article 10a(2) ETS-Directive, which provides that installations that 'are among the most efficient installations in the EU will in principle receive all allowances they need.'[137] The use of the correction factor, in contrast, penalises highly efficient energy recovery.

5.1.5 Compliance With Article 10a(2) ETS-Directive

In this context, the European steel industry association *Eurofer's* claims that Article 10a(2) ETS-Directive in fact 'obliges the Commission to set benchmarks "at the average performance of 10% most efficient installations in a sector"'[138]. If this interpretation was upheld by the Court of Justice, the approach might be seen to constitute and infringement of the Directive.

The language of the provision referred to, however, is vague and speaks of best performance to be the 'starting point' in defining benchmarks only; it does not contain a clear-cut obligation to design the fall-back methods in a way that, by all means,

[137] DG-CLIMA, 'Benchmarks for free allocation' (*European Commission*, 12 April 2012) <http://ec.europa.eu/clima/policies/ets/benchmarking/index_en.htm> accessed 18 August 2012.
[138] Eurofer, 'Court case of European steelmakers on benchmarks under the ETS to be judged on national level' (*Eurofer*, 8 June 2012) < http://www.eurofer.eu/index.php/eng/News-Media/Press-Releases/Court-case-of-European-steelmakers-on-benchmarks-under-the-ETS-to-be-judged-on-national-level2> accessed 10 July 2012.

accords full allocation to the best performers in a given sector. While this phrasing does not allow the Commission to design allocation methods that are completely arbitrary, it does concede a wide margin of appreciation. The chosen allocation method might be in excess of that margin if the rules cause a distortion of the respective markets (see below), but the mere deviation from the 'starting point' does not constitute an infringement on its own.

5.1.6 Conclusion on Allocation Deficit

The approach chosen to allocate allowances to installations generating waste heat electricity is a stopgap compensating for the failure of the ETS-Directive to take account of waste heat recovery, and for the failure of the Benchmarking Decision to address the problem more directly. Although it encourages energy recovery by re-interpreting occurring emissions as type (c) emissions, high efficiency would theoretically require an adjustment of correction factor ŋ and, that way, effectively penalise more efficient recovery. Full allocation, meanwhile, is impossible. The chosen approach incorporates a highly sophisticated technical formula, but as the values inserted are perfectly random, the formula—as applied to waste heat electricity—does not reflect due regulatory diligence, but is mere window-dressing. These problems do not indicate an excess of the Commission's margin of appreciation conceded by Article 10a(2) ETS-Directive, nor an infringement of the exception made for waste gas electricity in Article 10a(1)(3) from the rule of no allocation for electricity. But they raise the question whether the end justifies the means, or if it fails to deliver a fair competition environment. This is subject matter of the next subsection.

5.2 Level Playing Field

While the chosen allocation method might be arbitrary, it does not necessarily follow that the concerned industries are disadvantaged. However, the rules on free allocation for energy recovery are informed by, and hence to be interpreted in the light of, Recital (23) of the Directive, according to which they 'should (…) avoid undue distortions of competition on the markets for electricity and heating'.[139] Moreover, all European policy is subject to the requirement of promoting equal conditions of competition.[140] Therefore, the following subsection will examine if the approach creates a level playing field between waste heat electricity and other market participants, namely providers of waste gas for heat generation and suppliers of renewable and conventional electricity.

5.2.1 Delivery of Heat From Industrial Processes

5.2.1.1 Different Allocation Approaches

One problem might be that the use of off-gases for the production of heat (process heat or district heating) is privileged over the use for electricity production. A disparate

[139] Directive 2009/29/EC of the European Parliament and of the Council of 23 April 2009 amending Directive 2003/87/EC so as to improve and extend the greenhouse gas emission allowance trading scheme of the Community [2009] OJ L140/63, rec (23).
[140] Agreement on the European Economic Area [1994] OJ L1/3, art 1(1).

treatment would distort the market to the detriment of operators located in rural areas with little demand for district heating or process heat, as they do not have a choice but to opt for energy recovery in the form of electricity generation.

The table below provides an overview of allocation methods in relation to the production and consumption of waste gases. It should be noted that the producing and the consuming sub-installations can be part of the same industrial facility.

Production / Recovery	Consumption	Type of consumption	Allocation for recovery	Allocation for
Outside system boundary of PBM	Inside system boundary of	Product BM	Formula	Product BM
	Outside system boundary of PBM	Measurable	Formula	Heat BM
		Non Measurable	Formula	Fuel BM
		Electricity	Formula	None

Table no. 1: Allocation approaches for energy recovery and the consumption of that energy.[141]

While the use of the energy content of the off-gas (through combustion or heat recovery) is always eligible for free allocation, the consumption of the released energy for electricity production is not, though the consumption of the resulting heat for other purposes is. According to the relevant Heat and Fuel Benchmarks, installations not covered by a product benchmark which are consuming heat for a purpose other than electricity generation are allocated 62.3 allowances per TJ of thermal energy or 56.1 allowances per TJ of chemical energy consumed; this allocation is additional to the allocation made to the type (c) process emissions sub-installation. In the reference scenario of a 75/25 mix, this compensates roughly for the number of allowances not allocated as type (c) process emissions due to the deduction of energy-equivalent natural gas emissions and the 0.97 factor applying for all allocations made according to the process emissions approach.[142] The result is full allocation for the combustion of waste gas for the generation of heat, in contrast to 91.14% allocation for waste heat/gas electricity.

5.2.1.2 The Inherent Carbon Price of Electricity

However, the disparate levels of free allocation might be compensated by the inherent carbon price of electricity not paid due to the reduced intake of grid energy. As stated before, the generation and on-site consumption of electricity reduces the amount of allowances the operator has to pay for through the grid power price. Calculated on the

[141] DG CLIMA, Guidance Document 8 (2011) 18 (highlights added, captions modified).

[142] Fuel Benchmark: $0.97 * \frac{10}{11} * 0.00177\text{TJ/t} * (621.71 - 56.1 * 0.667 + 56.1)\text{t CO2/TJ} = 0.9995$.
Heat Benchmark: $0.97 * \frac{10}{11} * 0.00177\text{TJ/t} * (621.71 - 56.1 * 0.667 + 62.3)\text{t CO2/TJ} = 1.0092$.

basis of the transfer factor suggested by the industry,[143] the replacement of 1.33 MWh of grid electricity by waste heat electricity effectively reduces the allowance bill by one allowance. Provided there are no long-term consumption agreements between the operator and their power supplier and the process design allows for a substitution of grid power by waste heat electricity, this reduction can be seen as replacing an equal amount of allowances allocated for free.

This is particularly true in cases where electricity generation is more efficient than in the 75/25 reference case, which is based on hypothetical default values. The Finnfjord waste heat turbine, for instance, can generate up to 340 GWh a year at a maximum CO_2 output of 360 kilotons, resulting in an emission factor of just above one tonne of CO_2 per MWh. This performance is far better than in the reference scenario,[144] but it is not rewarded by the allocation formula for type (c) process emissions. However, compared against the transfer factor of the Nordic energy mix,[145] the achieved electricity output means that for every ton of carbon dioxide emitted, the cost of 0.57 grid electricity allowances is avoided.[146]

In other words: Adding the avoided grid allowances to the free allocation level of 91.14% theoretically results in a reduction of Finnfjord's allowance bill by almost 150% (compared to the number of allowances required for the production process itself),[147] exceeding the level of full allocation achieved through the sale of heat by far.

5.2.1.3 Comparability

However, there are several flaws to that comparison. Firstly, counting avoided grid power allowances towards the allocation level is appropriate only where grid power is actually being consumed; installations with a 'green' power supply (or a power supply that is at least 'greener' than the grid mix) cannot profit from the avoidance of passed-through allowances, or only to a lesser extent. Secondly, long-term consumption agreements between operators and power suppliers can make it impossible to reduce the intake of grid power without penalties. In that case, the electricity generated from waste heat must be sold off in competition with other sources of electric power, the relevant conditions of which are explored in the next subsection. Thirdly, the substitution of grid power by electricity generated from waste heat on-site faces technical difficulties. Waste heat occurs only after the chemical reduction process has been set off by large amounts of electric energy; a certain intake of grid power is

[143] 750g CO_2/kWh, Alliance of Energy Intensive Industries, 'Realistic evaluation of the indirect cost effects in the EU Emissions Trading Scheme for the analysis of the risk of carbon leakage and for financial compensation' (2009) 3.

[144] In which it is $\left(1.77\ \mathrm{GJ/t\ CO2}\ *\ 0.55 * 0.667 * 100\ \mathrm{t} * \frac{10}{11} * \frac{1}{3.6}\right)^{-1} = 6.1\ \mathrm{t\ CO2/MWh}$.

[145] 0.6 t CO_2 / MWh, see Energy Norway, Federation of Norwegian Industries and Industri Energi, 'Carbon Price Transfer in Norway' (2011) 3.

[146] $\frac{\frac{340}{360}\ \mathrm{MWh/t\ CO2}}{0.6\ \mathrm{t\ CO2/MWh}} = 0.57$.

[147] 0.9114 + 0.57 = 1.4814.

necessary to maintain the process and can be merely supported by electricity generated on-site.

Therefore, it is only under the condition of an ideal market and a specific process design that the generation of electricity is effectively favoured over the combustion of waste gas for the production of heat. It seems likely that under realistic circumstances, undue distortion of the respective markets will be avoided or at least minimised.

5.2.2 Electric Power Market

Apart from heat delivery as an alternative form of heat recovery, waste heat electricity also competes with electricity from renewable energy sources and combustion of fossil[148] fuels. As regards their treatment under the ETS schemes, the playing field between waste heat, conventional and renewable electricity is determined by two factors: Unlike the generation of waste heat electricity and conventional electricity, the production of renewable energy does not require the surrender of allowances; and unlike conventional electricity, waste heat electricity is eligible for free allocation.

5.2.2.1 Electricity From Renewable Energy

The industry argues that one of the drawbacks of waste heat electricity is its dependence on other, carbon-intensive processes. This might entail economic disadvantages such as a reduced ability to react to fluctuations in public power demand, or the exposure to increasing coal prices; those economic factors, however, remain outside the scope of a legal inquiry.

In the context of the ETS, the stated necessity results in a factual differentiation between waste heat and renewable energy: While neither electricity production from renewable energy nor the transformation of industrial heat into electric energy is subject to an obligation to surrender allowances, the latter cannot take place without other processes for which operators do have to surrender allowances. At the efficiency level described above, the process needed to generate sufficient amounts of heat requires the surrender of 0.094 allowances for every MWh of electricity produced.[149] It also requires the intake of grid power to set off the reduction process, and hence the indirect payment of grid power allowances; these allowances, however, are not quantifiable and will not be taken into account in the further consideration. Yet the fact remains that the generation of waste heat electricity requires the surrender of allowances, while the generation of electricity from renewable energy does not.

As both forms of electricity production are ‘green’ in the sense that they do not cause any additional greenhouse gas emissions, this disparate treatment might be found unfair: The Fifth Environmental Action Programme, which informs the establishment

[148] Biomass is excluded, see ETS-Directive, Annex I(1).

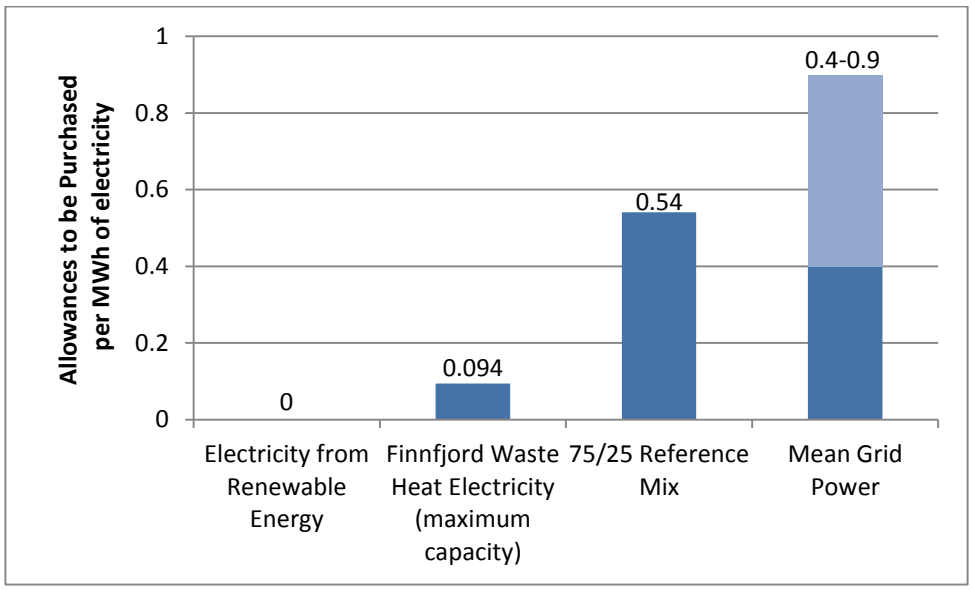
[149] $(1 - 0.9114) * \frac{18}{17} = 0.0938$.

of the trading scheme, requires the EU to promote *both* energy efficiency *and* the use of renewable energy sources.[150]

However, the focus on renewable and waste heat energy fails to take account of the ferro-alloy produced in the arc furnace: After all, it is the metal production that allowances have to be surrendered for, not the power production. It follows that, in terms of allowance allocation, there is no need for a level playing field between renewable electricity and waste heat electricity, as they are simply not playing in the same game. Incentives for green solutions as envisaged by the Fifth Environmental Action Programme can be created by way of more specific schemes such as state aid or special feed-in tariffs, provided that they are not incompatible with the objectives of the EEA Agreement.[151] But it is not up to the allocation scheme to balance out two processes that are related only through the common denominator of the (by-) production of 'green' electricity.

5.2.2.2 Mean Grid Power

A comparison between waste heat electricity and conventional grid power faces, in principle the same objection. However, the 2009 amendment explicitly stipulates that free allocation should avoid undue distortion of the respective markets; a greater competitive relation between grid electricity and waste heat electricity is at least assumed. In this context, however, the conditions provided by the ETS even seem to favour waste heat electricity over the electricity mix from the power grids: According to the transfer factors identified above, average suppliers of 'mean' electricity have to purchase 0.4-0.9 allowances for each MWh of electricity sold, while waste heat electricity requires the purchase of only 0.094 allowances.



[150] Decision No 2179/98/EC of the European Parliament and of the Council of 24 September 1998 on the review of the European Community programme of policy and action in relation to the environment and sustainable development 'Towards sustainability' [1998] OJ L275/1, art 3(2)(a).
[151] EEA Agreement (n 140), art 61(1),(3).

Diagram no. 9: Allowances Purchased by Method of Generation

The formula for type (c) process emissions is designed to minimise this discrepancy by subtracting hypothetical reference emissions from the historical activity level of the installation. However, the surprisingly low emission factor of waste heat electricity suggests that the energy content implied by the default value of a 75/25 mix might need to be adjusted in favour of a higher hypothetical CO-share, ultimately decreasing the historical activity level eligible for free allocation and increasing the inherent allowance bill of waste heat electricity.

On the other hand, the competitiveness of the primary product, which the allowances are allocated and must be surrendered for, must not be impaired by considerations of a potential market distortion caused by waste heat electricity, if this distortion is only of minor relevance. Moreover, it should be kept in mind that a non-quantifiable amount of grid electricity needs to be taken in first so as to trigger and maintain the reduction process, effectively increasing the allowance bill for waste heat electricity. Moreover, it should be noted that the number could change dramatically if the production of ferro-alloys lost the status of being exposed to a significant risk of carbon leakage and accorded the regular CLEF.

5.2.3 Conclusion on the Delivery of a Level Playing Field

The comparison of the relevant allocation levels produced the following results: Electricity from industrial waste heat receives fewer allowances than the combustion of waste gas for the generation of heat, but it also reduces the allowance bill rolled over by power suppliers. The number of allowances the operator has to purchase is higher than in the case of renewable energies, but lower than in the case of mean grid power energy.

The rules on free allocation have a potential of distorting the respective markets only in relation to the latter. Here, the subtraction of reference emissions from the historical activity level fails to balance the disparate allowance bills of waste heat electricity and grid electricity. Therefore, the default value for the off-gas mix (and hence, for the energy content of the off-stream) could be adjusted. Beforehand, however, the actual market distortion caused by this difference needs to be evaluated by an economic analysis, as the adjustment would primarily reduce the allocation level for the ferro-alloy production and impair its global competitiveness. This must not be made without prior comprehensive deliberation.

5.2.4 General Conclusion

Electricity generation from industrial waste heat adds significantly to the efficiency of the carbon-intensive process of ferro-alloy production; for that reason, the installation of corresponding turbines should be in the interest of both operators and policy-makers. The rules governing the free allocation of emission allowances fail to provide explicit incentives to install respective capacities through appropriate product benchmarks or

fall-back methods, but the industry was able to negotiate an interpretation of the existing provisions that, in case of heat recovery, leads to the free allocation of 91.14% of all allowances needed for the primary production process. The result, however, is a non-transparent nexus of binding provisions and nonbinding supplements that is based on arbitrary default values dressed in a pseudo-scientific formula.

While operators of installations generating industrial waste heat will never receive full allocation for their primary processes, the production of allowance-free electricity offsets this disadvantage (at least in a strictly legal context) and even favours waste heat electricity over grid power. It is tale-telling that an instant solution to this potential distortion would be to adjust the default value for the off-gas mix in favour of a higher CO-share; in fact, recent measurements performed by the two largest Norwegian ferro-alloy producers indicate that their off-gas consists of 100% CO_2. The perverse situation that the market distortion could be reduced by fudging an already hypothetical formula even more demonstrates how ill-suited the type (c) process emissions approach is to allocate allowances to installations generating electricity from industrial waste heat.

To redress the situation, the Commission needs to develop product benchmarks for ferro-alloys. It is no excuse to say that no sufficient data was available to do so, as the same is true for coke and hot metal, which the Commission managed to issue product benchmarks for 'based on information on relevant energy flows provided by the relevant BREF'[152]. The same could and should have been done in respect to ferro-alloys.

Hence, our answer to the research question is: Indeed, the political pressure applied by the industry has moved electricity from industrial waste heat out of the blind spot of Europe's energy policies, but the current situation is still far from perfect. Now, it is the turn of the Commission to develop product benchmarks for ferro-alloys that take into account the possibility of on-site consumption of electricity generated from the heat of the off-gas stream, and correct the regulatory hotchpotch the situation at hand represents.

[152] Benchmarking Decision (n 5), rec (11).

A COMPARATIVE STUDY OF THE OLD AND NEW TRACING AND LABELLING REQUIREMENTS FOR GMOS: EACH CONSUMER TO HIS OWN?

CHARLOTTE FERNANDO-MACVEAN[*]

This paper argues that the development of the EU's regulation of Genetically Modified Organisms (GMOs) has been successful. The 2003 reforms bring about a regime that satisfactorily balances the interests of both consumers and producers. Not only is consumer protection better delivered by the 'one door, one key' principle, but GMO regulation has become far more harmonised and efficient. The 2003 reforms also recognise, and give effect to, the economic importance of GMO development. The balance achieved is to be praised.

1 THE EVOLUTION OF THE LAW

A Genetically Modified Organism (GMO) is any organism, except humans, 'in which the genetic material has been altered in a way that does not occur naturally'.[1] An EU-wide memorandum on the marketing of GMOs existed from 1997-2004. Authorisation of GMOs within the EU was governed by Regulation 258/97,[2] but now falls under Directive 2001/18/EC,[3] as amended by Regulation 1829/2003.[4] Regulations 1829/2003 and 1830/2003[5] control the labelling and tracing of GMOs. 'GMOs are living organisms and… their flow cannot be turned off'.[6] Authorisation is therefore necessary. Public attitudes towards GMOs 'range from caution and doubt… to hostility and rejection'.[7] Accordingly, people must, by way of labelling and tracing, be able to 'follow their own preferences'[8] when choosing whether or not to consume GMOs. The 2003 amendments achieve the EU's aim of appropriately balancing the interests of consumers against the need to advance GM technology.

---

[*] Charlotte Fernando-MacVean, Newcastle Law School, Law LLB Stage Three.

[1] Council Directive 2001/18/EC of 12 March 2001 on the deliberate release into the environment of genetically modified organisms and repealing Council Directive 90/220/EEC [2001] OJ L106/1.

[2] Council Regulation (EC) 258/97 concerning novel foods and novel food ingredients [1997] OJ L043/1.

[3] Directive 2001/18/EC (n 1).

[4] Council Regulation (EC) 1829/2003 of 22 September 2003 on genetically modified food and feed [2003] OJ L268/1.

[5] Council Regulation (EC) 1830/2003 of 22 September 2003 concerning the traceability and labelling of genetically modified organisms and the traceability of food and feed products produced from genetically modified organisms and amending Directive 2001/18/EC [2003] OJ L268/24.

[6] David Hughes, Tim Jewell, Jason Lowther, Neil Parpworth and Paula de Prez, *Environmental Law* (4th edn, OUP 2002) 355.

[7] 'UK Parliamentary report show overwhelming public concern over GM crops' (28 September 2013) <http://www.non-gm-farmers.com/news_print.asp?ID=718> accessed 28 March 2013.

[8] Marine Fraint-Perrot, 'The European Union Regulatory Regime for Genetically Modified Organisms and its Integration into Community Food Law and Policy' in Luc Bodiguel and Michael Cardwell, *The Regulation of Genetically Modified Organisms: Comparative Approaches* (OUP 2010) 89.

2 THE SUBSTANTIAL EQUIVALENCE LOOPHOLE

In accordance with the 'substantial equivalence' principal under Regulation 258/97,[9] a GM crop 'similar' to its non-GM counterpart did not require a full safety assessment, and the producer needed only to notify the EU of its release onto the market. 'The idea of substantial equivalence is highly controversial';[10] an assessment is vital as 'the long term consequences of GMO technology are still unknown…(and) many people are afraid of the possible dangers'.[11] In *Monsanto SAS v the French Ministry of Agriculture and Fisheries*[12] the European Court rejected the argument that substantial equivalence prevented states from effectively investigating GMOs. It held that the lack of safety assessment and the fact that the principal could 'be used solely in order to speed up… administrative action'[13] did not invalidate it. However, in terms of adequately monitoring an uncertain area, the Regulation was unsatisfactory.

3 HINDERING THE FREE MARKET OBJECTIVE?

The criticism of Regulation 258/97 was focused upon its 'safeguard clause'.[14] If evidence of a GMO risking human health or the environment arose post-authorisation, a state could restrict or prohibit that particular GMO. Directive 2001/18/EC also contains this clause.[15] Member states were in charge of the clause's application and, although a risk assessment was required, member states were also responsible for setting the assessment threshold. This meant that bans could be placed upon GMOs in accordance with national public opinion. Such restraints hindered the EU's free market, as it caused variations of GMO regulation across the Union.

4 PRECAUTION UNDER DIRECTIVE 2001/18/EC

Directive 2001/18/EC applies to GMOs deliberately released by a producer into the environment. Producers must submit a notification to the relevant national body, containing an environmental risk assessment,[16] to gain authorisation. This procedure addresses the concerns surrounding the 'fast-track'[17] route under Regulation 258/97. The Commission must state whether authorisation complies with the 'precautionary principal', which calls for caution where an action involves a scientifically uncertain risk. Sunstein argues that overreliance upon the precautionary principal is detrimental as it 'leads in no direction',[18] and does not guide decision-making. Nevertheless,

[9] Council Regulation (EC) 258/97 (n 2) Art 5.
[10] Maria Lee, *EU Regulation of GMOs: Law and Decision Making for a New Technology* (Edward Elgar Publishing 2008) 73.
[11] European Parliament, *Report on the proposal for a directive of the European parliament and of the council amending directive 2001/18/EC concerning the deliberate release into the environment of genetically modified organisms, as regards the implementing powers conferred on the commission* (A6-0292/2007) 1.
[12] C-58/10 to C-68/10, [2011] OJ C311/10.
[13] Naomi Salmon, 'What's 'Novel' about it? Substantial Equivalence, Precaution and Consumer Protection 1997-2004' (2005) 7 Environmental Law Review 138, 141.
[14] Regulation (EC) 258/97 (n 2) Art 12.
[15] Directive 2001/18/EC (n 1) Art 23.
[16] ibid, Arts 13(1), 13(2)(b).
[17] Salmon (n 13) 139.
[18] Cass Sunstein, 'Beyond the Precautionary Principal' (2003) 151 University of Pennsylvania law

uncertainty surrounding the implications of GMOs justifies caution, and the need for a full-risk assessment strengthens the Directive in comparison to the 1997 Regulation.

5 CENTRALISATION OF THE RISK ASSESSMENT PROCESS

Regulation 1829/2003 amends the authorisation procedure for GM food, and foods containing, or produced from, GMOs. These foods must not pose a risk to humans, animals or the environment, mislead consumers, or nutritionally disadvantage consumers.[19] This detailed risk assessment means that the authorisation procedure is now far more rigorous than it was under Regulation 258/97. As with Directive 2001/18/EC, an application, including the risk assessment, must be made to the relevant national body. In contrast with the regime under the 2001 Directive, the national body must pass the application on to the European Food Safety Authority (EFSA) for authorisation. 'The Food and Feed Regulation reduces the role of the national authorities in risk assessment compared with the deliberate release Directive'.[20] This centralisation is beneficial to the EU free trade market; 'the more centralised process … [overcomes] the reluctance of the member states to accept each other's risk assessments, which made progress under the old law impossible'.[21] Technological advancement of GMOs enhances the economic strength of the Union; disagreements between states hinder this.

6 GMOS AND PUBLIC PARTICIPATION

A copy of the EFSA's opinion is released to the public, who can comment upon it. Although this public engagement appears to be an improvement upon previous legislation, there is scepticism surrounding its impact. Lee notes that 'the provision for public participation in the legislation is not particularly ambitious, [but is] limited to this opportunity to make comments'.[22] Moreover, a producer can withhold information from the public if it proves to the EFSA that the release would damage its competitive advantage. This limitation is unhelpful. Public mistrust focuses upon the motives of the GM industry, as well as its products.[23] Withholding information will not improve public confidence in the GMO sector, just as the 'fast-track' procedure under Regulation 258/97/EC did not.

7 ONE DOOR, ONE KEY

The 2003 Regulation's 'one door, one key principal' means that 'business operators may file a single application for the GMO and all its uses: a single risk assessment is performed and a single authorisation is granted'[24] covering GM food and feed. This

review 1003, 1054.

[19] Council Regulation (EC) 1829/2003, Art 4 (1).

[20] Lee (n 10) 66.

[21] Holder and Lee, *Environmental Protection, Law and Policy* (CUP 2007) 191.

[22] Lee (n 10) 82.

[23] ibid 79.

[24] Europa, 'GMOs in a Nutshell' <http://ec.europa.eu/food/food/biotechnology/qanda/GMO_qanda_2010_en.pdf> accessed 1 March 2012.

minimises the hindrance of the authorisation procedure compared with the 2001 Directive, under which applications for food and feed are made separately. 'The single authorisation procedure is conducted entirely at community level',[25] which improves harmonisation. The speed of the authorisation procedure under Regulation 258/97 benefitted GMO producers; the single application under Regulation 1829/2003 promotes this same speed, whilst also ensuring consumer safety.

8 THE GLOBAL FOOD CRISIS AND REGULATION OF GMOS

Article 34 of Regulation 1829/2003[26] governs a safeguard clause, as seen in the previous regime. However, the new clause can only be adopted at EU level. The EFSA is likely to have a greater workload than a national body. Therefore, the process under Article 34 is likely to be slower. Beddington argues that GMOs should not be rejected solely on the basis of ethical or moral concerns.[27] Allowing member states to enforce the safeguard clause opens up the possibility of states banning GMOs in line with such concerns of its citizens. 'In light of the magnitude of the challenges for food security in the coming decades',[28] GMOs are essential in alleviating the danger of a global food crisis. Hughes agrees that 'GM technology holds the answer to famine'.[29] The slower process is justified as it allows for neutral decisions to be taken at EU level.

9 A BALANCING OF INTERESTS

Regulation 1829/2003 places an obligation upon producers to update the Commission if new scientific evidence comes to light regarding the product's safety.[30] This ensures consumers' continued protection.

The authorisation of GMOs under Regulation 1829/2003 balances the interests of both consumers and GMO producers. The single-step procedure minimises the burden upon producers and the Union. The risk assessment is more rigorous than the one under Regulation 258/97, thus further promoting caution. The stricter procedure also benefits producers, as 'unless the regulatory framework is seen to be demanding and rigorous, advocates of the GMO technology will need to battle even harder to win over public opinion'.[31] The 1997 Regulation's controversial safeguard clause is now controlled by the dispassionate Union. 'Faced with endemic scientific uncertainty (and) mistrustful publics… the EU attempts to steer a course through deeply divided views'[32] and, as has been demonstrated above, does so appropriately.

---

[25] Fraint-Perrot (n 8) 90.
[26] Regulation (EC) No 1829/2003 (n 4) Art 34.
[27] Government Office for Science, *The Future of Food and Farming: Challenges and Choices for Global Sustainability* <http://www.bis.gov.uk/assets/foresight/docs/food-and-farming/11-546-future-of-food-and-farming-report.pdf> accessed 25 February 2012.
[28] ibid.
[29] Hughes (n 6) 353.
[30] Council Regulation (EC) 1829/2003 (n 4) Arts 9(3), 21(3).
[31] ibid 354.
[32] Lee (n 10) 103.

Nevertheless, 'it seems unlikely that the 'consumer interests' rationale will feed into the question of whether or not a GMO should be authorised'.[33] In terms of respecting consumer autonomy, an ability to trace the origins of GM foods and identify a GMO by its label is most important. 'One of the unspoken values of the regulatory framework for GMOs is the assumption that the market is the best place for the exercise of choice';[34] GMOs are necessary for economic advancement, but consumers within the market must be able to choose whether or not to consume them.

Under Regulation 258/97, both the labelling and tracing requirements were insubstantial. A GM crop fulfilling the requirements of the 'substantial equivalence' test did not require labelling. The flaws of the test have been discussed previously in this paper and, in the area of labelling, meant that consumers wanting to 'follow their preferences when choosing what to eat'[35] were faced with an obvious hurdle.

Regulation 1829/2003 requires that all food be labelled as to whether or not it contains, or is produced from, GMOs.[36] 'Within the context of a regulatory field plagued by uncertainty and controversy, it seems entirely rational and in line with the precautionary approach to food and safety… to remove the contentious and artificial concept of substantial equivalence'.[37] Article 12 allows consumers to make an informed decision on whether to consume GMOs.

10 AN UNKNOWN PRESENCE?

Under Regulation 1829/2003, the labelling requirement does not apply to those products made up of 0.9% GMOs or less, 'provided that this presence is adventitious or technically unavoidable'.[38] 'A central ambiguity resides in the fact that consumers do not realistically have access to food that is guaranteed free from GMOs'.[39] Seemingly, this loophole inhibits consumer ability to make educated decisions regarding consumption of GMOs, as it was the case with 'substantially equivalent' products under Regulation 258/97/EC.

The threshold actually 'amounts to a confession that it is impossible to prevent cross-contamination by GMOs'.[40] In accordance with the precautionary principal, GM innovation is cautious. However, natural cross-pollination makes a 0% threshold impossible.[41] Consequently, as 'it is not possible to prove scientifically'[42] that a foodstuff is free from GMOs, a lower threshold would create an unrealistic burden on producers. In comparison to countries such as Japan and Canada, where the threshold

[33] ibid 87.
[34] ibid 107.
[35] Fraint-Perrot (n 8).
[36] Regulation (EC) No 1829/2003 (n 4) Art 12.
[37] Salmon (n 13) 144.
[38] Regulation (EC) No 1829/2003 (n 4) Art 21.
[39] Fraint-Perrot (n 8) 94.
[40] ibid 93.
[41] Case C-442/09 *Karl Heinz Bablok and Others v Freistaat Bayern* [2011] OJ C442/09.
[42] Case T-13/99 *Pfizer Animal Health SA v Council* [2002] ECR II-3305, 145.

is 5%,[43] the EU can be seen to provide consumers with a significantly greater level of protection.

Food products obtained from animals fed on GM feed do not need to be authorised or labelled.[44] This limitation significantly restricts consumers' ability to exercise preference over whether or not they consume GMOs. Furthermore, the problem will worsen if GMO popularity grows, and farmers are enticed by the benefits of genetically enhanced feeds.

These limitations upon consumer autonomy are justified in the interest of advancing GM technology. 'Within the EU, there are concerns over a slowdown in biotech crop research and the long run international competitiveness of EU agriculture'.[45] The 2003 Regulation strengthens the position of consumers in comparison to the previous legislation, whilst still recognising the need to encourage sector growth. 'No laws specific to biotechnology products'[46] exist in countries such as the US. Therefore, overregulation of EU GMOs would be detrimental to trade. If other countries are actively advancing and benefitting from GM technology, EU economic competitiveness will be harmed unless it follows suit. This rationale also argues against the reform proposal to allow states to 'opt out' of producing GM crops on the grounds of, for example, socio-economic considerations.[47] 'There are many powerful reasons for maintaining the potential of GMOs'.[48] The current Regulation allows for consumer choice so far as is practical, whilst simultaneously recognising the benefits that GMOs may provide.

Foods produced from, consisting of, or containing GMOs must be traceable 'at all stages of their placing on the market'.[49] Operators dealing with such products must hold information on the GMOs, including the unique identifier for that GMO, for a period of 5 years. Difficulties arise in cases where the GM presence is miniscule. To trace such a presence places a huge burden upon producers. However, as 'scientists disagree about the risks and net benefits'[50] of GMOs, traceability is essential. GMOs must be traceable

[43] Colin Carter and Guillaume Gruere, 'International Approaches to the Labelling of Genetically Modified Foods' (*Agricultural Marketing Research Centre*, March 2003) <http://www.agmrc.org/media/cms/cartergruere_929BEB69BA4EE.pdf> accessed 20 February 2012.

[44] Regulation (EC) 1830/2003 (n 5) Recital 16.

[45] Carter and Gruere (n 43).

[46] Guide to U.S. Regulation of Genetically Modified Food and Agricultural Biotechnology Products (*Pew Initiative on Food and Biotechnology)* <http://www.pewtrusts.org/uploadedFiles/wwwpewtrustsorg/Reports/Food_and_Biotechnology/hhs_biotech_0901.pdf> accessed 25 February 2012.

[47] Commission, 'Proposal for a Regulation of the European Parliament and of the Council amending Directive 2001/18/EC as regards the possibility for the Member States to restrict or prohibit the cultivation of GMOs in their territory' COM (2010) 380 final.

[48] Lee (n 10) 22.

[49] Regulation (EC) 1830/2003 of Genetically Modified Organisms and the Traceability of Food and Feed Products produced from Genetically Modified Organisms, Art 3(3).

[50] Margaret Grossman, 'Traceability and Labelling of Genetically Modified Crops, Food, and Feed in the European Union' (2005) 1 Journal of Food Law and Policy 43, 44.

so that the authorities can remove them from the market, and protect consumers in instances where safety concerns arise. Traceability also allows consumers to discover the history of their food and thereby follow their preferences when choosing whether or not to consume GMOs.

11 CONCLUSION

Regulation 1829/2003 successfully builds upon Regulation 258/97/EC and Directive 2001/18/EC. Success in this context is measured against the appropriate balance between the interests of protecting consumers and the interests of not unduly inhibiting the advancement of GMO technology. The 'artificial' substantial equivalence test is gone, meaning that the authorisation procedure is more rigorous than before, and better protects consumers. Safeguard clauses can only be adopted at EU level, meaning that the advancement of GM technology can be neutrally governed, and GM regulation is harmonised. The one-step authorisation procedure relieves the burden on GM producers, and makes the procedure more efficient. The 1829/2003 labelling requirements appropriately balance the need for consumers to be informed with the need for GM technology to develop. The 2003 tracing requirements protect consumers in cases of hazardous GMOs, and allow consumers to trace the products that they are consuming, thereby facilitating informed decisions.

ACKNOWLEDGEMENTS

We would like to thank the following organisations for their kind support and generous contributions which have been invaluable to the production of this journal.