



THE
OXFORD UNIVERSITY
UNDERGRADUATE LAW
JOURNAL

~

ISSUE TEN
TRINITY TERM
2021

The Editorial Board wishes to thank our platinum sponsors—3 Verulam Buildings, Pump Court Tax Chambers, Serle Court Chambers, and South Square Chambers—for their generous sponsorship of the Journal.



PUMP COURT TAX CHAMBERS



serle court



The views expressed by the contributors are not necessarily those of the Editorial or Honorary Board of the Oxford University Undergraduate Law Journal. Whilst every effort has been made to ensure that the information contained in this journal is correct, the Editors and the authors cannot accept any responsibility for any errors or omissions, or for any consequences resulting therefrom.

© 2021 Individual authors



THE TENTH EDITORIAL BOARD**EDITOR-IN-CHIEF**

Bruno Ligas-Rucinski
Christ Church (Graduated 2020)

EDITOR

Ben Coady
Lincoln College

VICE-EDITORS

Richard Wagenländer
Somerville College

Merit Flügler
Merton College

Kee Hwan Yeo
Lincoln College

SENIOR ASSOCIATE EDITORS

Amy Hemsworth
St Edmund Hall

Kathleen Wang
Pembroke College

ASSOCIATE & BLOG EDITORS

Kacper Kryk (A)
Corpus Christi College

Arista Lai (A)
Trinity College

Rachel Durnford (A)
Magdalen College

Zaynab Sarguroh (A)
St Hugh's College

Eve Thomson (B)
St Catherine's College

Mihir Rajamane (B)
Magdalen College

Caitlin McArthur (B)
St Edmund Hall

Chris Lippert (B)
Merton College

Dolly Xu (B)
Merton College

Henry Fahrenkamp (B)
Magdalen College

Weronika Galka (B)
St Peter's College

Cai Fangwei (B)
Worcester College

SECRETARY

Daria Koukoleva
Christ Church

TREASURER

Zaynab Sarguroh
St Hugh's College

THE HONORARY BOARD

Sir Nicolas Bratza

Professor Michael Bridge

Donald Findlay QC

Professor Christopher Forsyth

Ian Gatt QC

The Rt Hon. the Lord Judge

Michael Mansfield QC

The Rt Hon. the Lord Neuberger of Abbotsbury

The Rt Hon. the Lord Phillips of Worth Matravers

The Lord Pannick QC

His Honour Judge Gordon Risius CB

Dinah Rose QC

Sir Konrad Schiemann

The Rt Hon. the Lord Wilson of Culworth



CONTENTS

I	Introduction to the Edition Bruno Ligas-Rucinski: Editor-in-Chief	8
II	Foreword (Public Law) The Rt. Hon. Lord Wilson of Culworth: Former Justice of the Supreme Court of the United Kingdom and OUULJ Honorary Board Member	11
III	Foreword (Private Law) The Rt. Hon. Lord Neuberger of Abbotsbury: Former President of the Supreme Court of the United Kingdom and OUULJ Honorary Board Member	16
IV	Prizes	20

PUBLIC LAW ARTICLES

V	Regulating the Robot: A Toolkit for Public Sector Automated Decision- Making (Part A) Benjamin Cartwright	23
VI	Bias and Discrimination in Opaque Automated Individual Risk Assessment Systems (Part B) Gianna Seglias	53
VII	Male Mothers: The Law's Struggle to Capture the Modern Family Caragh Deery	80

VIII	The Future of the Energy Charter Treaty in a post-<i>Achmea</i> Era Sze Hian Ng	111
IX	‘Gentlemen at home, hoodlums elsewhere’: The Extraterritorial Application of the ECHR Petra Stojnić	137
X	<i>R (Dolan) v Secretary of State for Health and Social Care: Legality in the Time of Coronavirus</i> John Yap & Nicholas U Jin	171

PRIVATE LAW ARTICLES

XI	The Public Interest in Private Law: Re-Calibrating the Law of Tort Juana De Leon	212
XII	Overcoming the Law’s Marriage to the Past Merit Flügler	239
XIII	Saviour Siblings, Commercial Organ Donation and Commercial Surrogacy: Finding the Right Balance Bruno Ligas-Rucinski	265

INTRODUCTION TO THE EDITION

Bruno Ligas-Rucinski
Editor-in-Chief

It is my great pleasure and privilege to introduce the 10th Edition of the Oxford University Undergraduate Law Journal (OUULJ).

This year's edition of the Journal is a particularly special one for several reasons. First, the Journal celebrates its 10th anniversary. Founded in recognition of a boundless appetite for intellectual exploration amongst students, the Journal has since been providing a platform for interesting ideas to be debated and advanced by the very best across the country. This year, we received the highest number of submissions ever. Clearly, the desire to engage with challenging issues shows no sign of waning.

Secondly, we are delighted that two former justices of the Supreme Court of the United Kingdom accepted our invitation to judge the best private and public law submissions. I thank the Rt. Hon. Lord Neuberger of Abbotsbury and the Rt. Hon. Lord Wilson of Culworth for their interesting, detailed and touching forewords. Both judges emphasised the thought-provoking nature of the articles and the difficulty they had in selecting the winning submissions. All the authors are to be applauded for the quality and significance of their contributions to public and private law literature.

Thirdly, all of us here at the Journal are grateful to our sponsors, Allen & Overy, South Square Chambers, 3 Verulam Buildings, Serle Court Chambers and Pump Court Tax Chambers. Their generosity has provided us with the money to fund the prizes for the best private and public law submissions, to subsidise the printing of the Journal and to enable us to fund

the most innovative enterprise that we have embarked on this year, the Oxford Undergraduate Law Podcast.

The Podcast complements the work of the Journal and the Oxford University Undergraduate Law Blog (OUULB) by providing a platform for more topical subjects in the law. For example, the latest episodes include ‘Patenting Vaccines’ with Professor Justine Pila and ‘Regulating Big Tech’ with Professor Ariel Ezrachi. Looking ahead, we aim to use this digital platform to host sponsored events which might assist students in securing a coveted pupillage or training contract. Many thanks are owed to Bianca Dammholz and Siobhan Tan, our two Podcast Editors. Without their effort and drive, the endeavour simply would not have been possible.

Furthermore, we have looked not only within Oxford for inspiration, but beyond too. We have continued to share calls for submissions with our partner journals in Cambridge, Trinity College Dublin, the LSE, and Bristol to stimulate the flow of academic talent between these institutions. Moreover, we participated in a World Law Reviews Seminar. The event was hosted by the University of Bologna Law Review and was attended by the Cambridge Law Review and the Auckland Law Review. It was a fantastic opportunity to exchange ideas regarding how best to increase the reach of our journals, how to improve the editing process, and what events we would like to collaborate on in the future.

As a blind-reviewed, rigorously edited journal, the OUULJ could not function without the industry of its team of editors. The Oxford law degree leaves only a limited amount of time to pursue extra-curriculars. I am therefore tremendously grateful to our editors for their diligence in keeping to deadlines. They must also be praised for their insightful comments, which have been instrumental in elevating the quality submissions

within this journal even further. Special thanks must go to the Senior Associate Editors, Amy Hemsworth and Kathleen Wang, for their unwavering commitment. Thanks must also go to our treasurer, Zaynab Sarguroh, and secretary, Daria Koukoleva, who helped us effectively manage the administrative side of the Journal.

Finally, I owe heartfelt thanks to the Senior Editorial Team. Our Editor this year, Ben Coady, has proven to be the quintessential polymath. He has been an integral part of the sponsorship drive, in securing the contributions from our esteemed judges and in the day-to-day management of the Journal. This decennial edition of the OUULJ would not have been possible without his tenacity, wisdom and dedication. Complementing Ben's strengths, our three Vice-Editors, Richard Wagenländer, Merit Flügler and Kee Hwan Yeo, have worked admirably this year. They have each upheld the quality of the OUULB and supported the wider editorial team. I wish them all the very best of luck in their future endeavours.

To conclude, this edition marks the third and final year that I have worked on the Journal. In his touching foreword, Lord Neuberger highlights the enormous value that academic journals can give to the formation of young lawyers. I can attest to this wholeheartedly, as I finish my term as Editor-in-Chief more humbled by the brilliance of my peers than ever before. My lasting hope is that you, the reader, come away from this 10th edition of the Journal inspired, informed, and excited for the future of this publication.

FOREWORD (PUBLIC LAW)

The Rt. Hon. Lord Wilson of Culworth
Former Justice of the Supreme Court of the United Kingdom
OUULJ Honorary Board Member

The most important recent decision of the UK Supreme Court in the realm of public law is surely its rejection of the application for judicial review of the decision to refuse leave to the applicant to enter the UK in order to participate in her appeal against the order which deprives her of UK citizenship: *R (Begum) v Secretary of State for the Home Department*.¹ As always, the judgment of Lord Reed appears carefully measured. The Home Secretary's decision had been based on asserted concerns that the effect of Ms. Begum's return to the UK would be to endanger national security; and clearly the court had to afford respect to them. But, unless they were a trump card, how much respect? The court concluded that her application should be stayed until she could participate effectively in the appeal without endangering national security. It conceded that this solution was imperfect in that the deprivation of her citizenship (which has generated widespread public controversy) would therefore remain in force, incapable of challenge for a period upon which the court clearly felt unable to set a date even for review. The question is whether this undeniably unattractive result should have affected the amount of respect which fell to be paid to the asserted concerns. My experience is that the judges of the Supreme Court strive to their utmost to prevent their personal predilections from playing a part in the evaluative exercises required of them. But do they always succeed? Can we imagine, for example, that Lord Kerr, whose recent death is so widely

¹ [2021] UKSC 7, [2021] 2 WLR 556.

mourned, would have associated himself with the court's evaluation?

A different but linked question is whether we should expect judges entirely to exclude consideration of the political and social consequences of deciding a case in one way rather than in another. It is prompted by the hard-hitting article by John Yap and Nicholas Jin in this edition on the Court of Appeal's rejection of the challenge to the validity of the regulations by which in March 2020 the Government introduced into England the severe lockdown measures designed to combat the Covid-19 pandemic: *R (Dolan) v Secretary of State for Health*.² John and Nicholas charge the court with excessive deference to executive decisions made in time of emergency. In fact the regulations had been repealed during the hearing at first instance but by no means all of the political and social consequences of any decision to declare them invalid had then evaporated.

I am reminded of my own participation in an appeal to the Supreme Court which had potentially substantial political and social consequences: *Moohan v The Lord Advocate*.³ The appeal, heard on 24 July 2014, challenged the lawfulness of legislation which precluded all convicted prisoners in Scotland from voting in the referendum on Scottish independence fixed to take place only eight weeks later. It was doubtful whether, were the legislation to be held unlawful, the necessary remedial measures could be brought into effect within that time. At an early stage, therefore, the profound consequences of upholding the appeal troubled me. Nevertheless, by the end of the hearing, I had become convinced that the blanket exclusion of all the prisoners from the franchise breached the human rights of at any rate some

² [2020] EWCA Civ 1605, [2021] 1 WLR 2326.

³ [2014] UKSC 67, [2015] AC 901.

of them. It soon transpired, however, that a majority of my colleagues proposed to dismiss the appeal and that my judgment would be one of dissent. I confess that at that point my concern at what I considered to be the unlawful exclusion of prisoners from the franchise was tempered by relief that, by my judgment, I would not be contributing to substantial political and social upheaval; and my question to you undergraduates is whether these sentiments of mine were professionally inappropriate.

In her nicely provocative article Petra Stojnic touches on another subject of acute current relevance—the subjection of UK armed forces to obligations to uphold the human rights of those with whom they are in armed conflict abroad. Her focus is the decision of the Strasbourg court in *Al-Skeini v UK*.⁴ Then in an era of expansive disposition, the court held that the UK breached the rights of non-detained Iraqi civilians whom its forces shot in the course of a military operation; for, although the UK did not control that territory, it assumed responsibility for the maintenance of security there, with the result that the civilians were held to fall within its jurisdiction for Convention purposes. The burden of Petra’s article is that the court in *Al-Skeini* did not go far enough and that, if the UK caused what would, in the event of jurisdiction, be a breach of the Convention, then jurisdiction should depend on an assessment of the scope of its responsibility on the facts of each case. She might have been wise to refer to, if only to disagree with, our extensive review of *Al-Skeini* in *Mohammed v Secretary of State for Defence*,⁵ but the real questions—apt for answers in next year’s edition?—are whether Petra’s solution would be appropriate as a pan-European jurisdictional criterion or whether it is practicable either to expect soldiers in the heat of armed conflict so to confine their activities as to

⁴ [2011] 7 WLUK 207, (2011) 53 EHRR 18.

⁵ [2017] UKSC 2, [2017] AC 821.

respect the human rights of those ranged against them or indeed to expect courts to be able satisfactorily to judge whether soldiers in that situation had infringed those rights.

In her article Caragh Deery considers the decision of the Court of Appeal in *R (McConnell and YY) v Registrar General for England and Wales*,⁶ which addressed yet a third subject of extreme topicality. Shortly after he had secured recognition as transgender, Mr McConnell became pregnant and gave birth to YY. In the proceedings he unsuccessfully challenged his description as YY's 'mother' in the Register of Births. The court held that the Registrar was required to reflect the biological reality of the birth, which was that Mr McConnell was the mother and therefore that, notwithstanding his legal status as male, he could not instead be recorded as the father. Caragh's examination of the case is well-balanced and attuned to the surrounding sensitivities; and while, in the end, she endorses the decision as in accordance with the present law, she calls for discussion whether, without forfeiting biological accuracy, other descriptions of parenthood might be devised which would better serve our differing modern identities.

Then we find the article by Sze Hian Ng on the effect of the decision of the Luxembourg court in *Slovak Republic v Achmea BV*.⁷ A bilateral investment treaty between Holland and Slovakia provided for disputes to be resolved by an arbitral tribunal; and the court held the provision to be unlawful because the tribunal would be interpreting EU law while lying outside the network of judicial bodies solely authorised to interpret it under EU treaties. The question is whether the effect of the decision prohibits access by EU members to arbitral tribunals established under the important Energy Charter Treaty, to which there are many other

⁶ [2020] EWCA Civ 559, [2021] Fam 77.

⁷ [2018] 4 WLR 87.

contracting parties worldwide. Sze Hian's answer is 'perhaps not'. I hope that he becomes an advocate for he has a rare forensic skill: he makes a dry subject sound interesting.

It remains only to refer to two complementary articles on the need to control the use of algorithms in automated decision-making by public authorities. The authors recognise the benefits of such systems in terms of consistency but address the risks that some such decisions will be infected by unlawfulness which will be difficult for the victims to establish. In his article Benjamin Cartwright suggests a suite of controls to operate before the algorithm is set to work, in particular its scrutiny by a specialist body, an Algorithm Commission. In hers Gianna Seglias addresses the particular use of algorithms in the assessment of the risk of a person's future unlawful behaviour in deciding whether to bring a charge or to grant parole. She suggests that a muscular use of the laws against discrimination (the principles of which I always found fiendishly difficult to apply) would help to establish the unlawfulness of the decision in court. These thoughtful articles deserve high-level professional attention.

For the purpose of writing this foreword I have read each of these articles twice. Instead of being a burden, it was a pleasure. For, in different ways, each is excellent. Congratulations!

June 2021

FOREWORD (PRIVATE LAW)

The Rt. Hon. Lord Neuberger of Abbotsbury
Former President of the Supreme Court of the United Kingdom
OUULJ Honorary Board Member

The relationship between academic lawyers and practising lawyers in this country has changed enormously over the past sixty years and beyond all recognition over the past 200 years—and the change had been very much for the better.

Among the quainter principles developed by judges was the rule that what was written in a law book or article, however distinguished, experienced and respected the author was, could not be cited in court or relied on in a judgment if the author was still alive. 200 years ago, the principle was cited in judgments in common law and equity cases as clear and well established.¹ Less than 100 years ago, in 1927, the rule was confirmed by the then-Master of the Rolls.² Indeed, five years later, the rule was confirmed by the Law Lords in the famous case of *Donoghue v Stevenson*,³ where it was said that ‘the work of living authors, however deservedly eminent, cannot be used as authority’, although it was somewhat grudgingly accepted that ‘the opinions they express may demand attention’. Even after an author had died, he (as it was almost always ‘he’) could not rely on being taken into account by judges. Thus, in 1814, Lord Chancellor Eldon said that ‘One who held no judicial situation could not regularly be mentioned as an authority’.⁴ (Perhaps consistently with that

¹ *Turner v Reynard* (1819) 1 J & W 39, 44; *Taylor v Curtis* (1816) 6 Taunt 608.

² *Re Ryder & Steadman’s Contract* [1927] 2 Ch 62, 74.

³ *Donoghue v Stevenson* [1932] AC 562, 567.

⁴ *Jones v Jones* (1814) 3 Dow 1, 15.

approach, judges broke the ‘only read when dead’ rule from time to time when it suited them—especially when the author was a judicial colleague).⁵

Various lame, even laughable, reasons were advanced to justify this rule, but perhaps the most convincing was that, until the latter half of the 19th century, academic law could be (and was⁶) described as ‘a fairly moribund, amateurish profession’. However, with the growth in numbers and quality of academic lawyers and academic articles and books it became inevitable that judges would start to take academic works seriously. Interestingly, it began happening in the 1880s, just after Oxford and Cambridge had set up their law faculties. In 1882, the formidable Sir George Jessel MR said that, when faced with a controversial legal principle, it was always right to ask: ‘what do the text books say?’ as ‘although the text books do not make law they show more or less whether a principle has been generally accepted’.⁷

With the appearance of highly respected textbooks (like Megarry’s *Law of Real Property*, written with HWR Wade) and the development of highly reputable Journals (such as the *Law Quarterly Review*), the position has now changed completely. ‘[N]owadays judges read academic articles as part of their ordinary judicial activity’.⁸ If one had to single out one judge who was responsible for this, I think it would be Lord Goff of Chieveley. As a judge in the High Court, Court of Appeal and House of Lords, he frequently cited, paid tribute to, and relied on academic works. It is interesting to note that, in 1966 as a practising barrister (and in subsequent editions as judge) he wrote a seminal

⁵ *Cholmondeley v Clinton* (1820) 2 Jac & W 1, 151-152 (the same judge who decided *Turner* a year earlier).

⁶ Neil Duxbury, *Jurists and Judges: An Essay on Influence* (Hart 2001) 71.

⁷ *Henty v Wrey* (1882) 21 ChD 332, 348.

⁸ *Re OT* [2004] EWCA Civ 653 [43] (Longmore LJ).

book with a highly respected academic, the late Gareth Jones, on *The Law of Restitution*, now retitled *The Law of Unjust Enrichment*. In a speech given in 1987, Lord Goff said that it was ‘difficult to overstate the influence of the jurist in England today—both in the formation of young lawyers and in the development of the law’,⁹ a sentiment with which almost all judges and indeed almost all practising lawyers would agree.

The very salutary fact that judges are now taking proper notice of academic writing should not blind one to the differences between the attitudes of judges and of academic lawyers. One view as to the difference between academic and judicial writing was expressed in a judgment by Sir Robert Megarry, himself a great writer of legal books. In a 1969 case,¹⁰ disagreeing with what he had said as an author in one of his own books, he said:

The process of authorship is entirely different from that of judicial decision. The author, no doubt, has the benefit of a broad and comprehensive survey of his chosen subject as a whole, together with a lengthy period of gestation, and intermittent opportunities for reconsideration. But he . . . lacks the advantage of that impact and sharpening of focus which the detailed facts of a particular case bring to the judge. Above all, he has to form his ideas without the purifying ordeal of skilled argument on the specific facts of a contested case. Argued law is tough law.

The academic side of the argument got the last laugh, however, as less than ten years later, the Court of Appeal held

⁹ Lord Goff, ‘Judge, Jurist and Legislature’ (1987) 2 Denning Law Journal 79, 92.

¹⁰ *Cordell v Second Clanfield Properties Ltd* [1969] 2 Ch 9, 16-17.

that Megarry the author had been right and Megarry the judge had been wrong.¹¹

The important point is that judges can benefit very substantially from the perspective and thoughts brought by academic experts to a particular subject and the broad and rigorous examination to which they have subjected it. That perspective can often provoke ideas, which can be tested in court, but which would not otherwise have come to light in proceedings. In that way we ensure that, to borrow from Oliver Wendell Holmes, the law develops through experience in the widest possible way.

Each of the three impressive papers included in this edition of the OUULJ are excellent examples of the enormous value of academic articles can provide to the development of the law—and they also represent excellent examples of what Lord Goff referred to as the contribution made by Universities to ‘the formation of young lawyers’. I have little doubt that when the law of nuisance/defamation, the law of surrogacy/organ donation, or the law of spousal guarantees come to be considered by the courts or by the legislature, the impressively developed thoughts of Juana De Leon, Bruno Ligas-Rucinski and Merit Flügler respectively will receive considerable attention. I congratulate them and their supervisors on their thought-provoking, persuasive and well-expressed papers.

May 2021

One Essex Court Temple London EC4

¹¹ *St Edmundsbury and Ipswich Diocesan Board of Finance v Clark (No 2)* [1975] 1 WLR 468.

PRIZES

Best Public Law Submission to the Tenth Edition of the Oxford University Undergraduate Law Journal (2021):

Male Mothers: The Law's Struggle to Capture the Modern Family

Caragh Deery

St Hilda's College, Oxford

Lord Wilson made this comment:

'My background is in family law and it may seem that, in choosing an article on a point of family law, I have too easily gone back to my roots. But, albeit by what in racing used to be called a short head and is now called a nose, Caragh Deery's article is—objectively—the best.'

The Editors are grateful to Lord Wilson for adjudicating upon the public law articles.

Best Private Law Submission to the Tenth Edition of the Oxford University Undergraduate Law Journal (2021):

Saviour Siblings, Commercial Organ Donation and Commercial Surrogacy: Finding the Right Balance between Acceptable Instrumental Use and Impermissible Commodification

Bruno Ligas-Rucinski

Christ Church, Oxford

Lord Neuberger made this comment:

'I genuinely found it very difficult to choose between the three papers, as they are all first class in terms of topic, argument and presentation. On balance I would choose Saviour Siblings, Commercial Organ Donation and Commercial Surrogacy:

Finding the Right Balance between Acceptable Instrumental Use and Impermissible Commodification by Bruno Ligas-Rucinski. In making this choice, it is possible that I may have been influenced by the fact the other two papers were on property law topics: I was a property lawyer, and so may have been more easily impressed by a paper on a subject in a very different area.’

The Editors are grateful to Lord Neuberger for adjudicating upon the private law articles.

*The Editorial Board wishes to thank our prize sponsor—**Allen & Overy LLP**—for their generous contribution to the prize fund.*

ALLEN & OVERY

PUBLIC LAW ARTICLES

Regulating the Robot: A Toolkit for Public Sector Automated Decision-Making

Benjamin Cartwright*

Abstract—The use of automated decision-making (ADM) technologies by public sector bodies has developed significantly in recent years. As a result, decisions are increasingly made with limited, or no, human involvement. The need to regulate such technologies is therefore imperative. This article responds to the recent proposal of Lord Sales to establish an ‘Algorithm Commission’, to regulate and advise public bodies on ADM system design and scrutiny. While such a proposal is worthy, it is argued that a better method by which to regulate ADM technologies is through a toolkit of regulatory mechanisms, by deploying targeted impact assessments, codes of practice, technical tools and the existing data protection framework.

* BCL candidate, Harris Manchester College, University of Oxford. This paper was written during my LLB at University College London. I am grateful for the insightful guidance and comments on earlier drafts by Professor Rick Rawlings, Chris Moss and the OUULJ Editorial Board. All errors remain my own.

Introduction

‘The algorithms of the law must keep pace with new and emerging technologies’.¹ Nowhere is this truer than in the context of public sector decision-making, where the recent uptake of artificial intelligence (AI) based automated decision-making (ADM) systems has been astonishing. From use in tax audit, to improving health standards and policing, automated processes have served to transform the administrative state, using technology to raise standards, reduce error rates and cut costs. A recent US study indicates that ADM systems are now used in more than 45% of federal agencies, with many more such systems in the pipeline.² This increasing ubiquity, and the challenges associated with ADM systems, makes it essential to effectively regulate them: their deployment will only continue in the coming years.

Considering this, Lord Sales has proposed an ‘Algorithm Commission’ to regulate and advise Parliament, the courts and government, on matters of ADM system design, scrutiny and adjudication.³ In this article, it will be argued that while such a Commission may be helpful in ensuring accountability, regulatory *bodies* such as that proposed by Lord Sales are an inferior method of regulating ADM systems, when compared to instrumental legal and regulatory *mechanisms*, such as algorithmic impact

¹ R (*Bridges*) v *Chief Constable of South Wales Police* [2019] EWHC 2341 (Admin) [2020] 1 WLR 672 [1].

² David Freeman Engstrom, Daniel E Ho, Catherine M Sharkey and Mariano-Florentino Cuéllar, *Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies* (Administrative Conference of the United States 2020) 6.

³ Lord Sales, ‘Algorithms, Artificial Intelligence and the Law’ (Sir Henry Brooke Lecture for BAILII, 12 November 2019) <<https://www.supremecourt.uk/docs/speech-191112.pdf>> accessed 10 January 2020.

assessments, codes of practice, technical tools and the existing data protection framework. It is argued that a new legal toolkit, fashioned around such regulatory mechanisms, will be better placed to mitigate the accountability risks associated with ADM system development. Ultimately, intrinsic regulation of ADM systems, albeit necessarily supported by a regulatory body, is more likely to guarantee adherence to the overarching legal regime. It is therefore not argued that there should be *no* role for a regulatory body; *a* body will be needed to deploy the toolkit argued for, and to ensure that the instrumental mechanisms meet their intended purposes in ensuring algorithmic accountability. As governance is increasingly digitised, the need to modernise accountability mechanisms which public bodies may deploy only grows.

What follows takes a three-part structure. Part 1 conceptualises and briefly surveys various legal issues related to ADM. Part 2 focuses on the ambit of an algorithm regulator responsible for ensuring *ex-ante* accountability. Part 3 outlines the proposed legal toolkit which the regulator would be responsible for deploying and overseeing. The focus of the regulatory toolkit proposed will be on *ex-ante* regulation of ADM systems, given that hitherto, most discussion of ADM system regulation has been to ensure *ex-post* (judicial) accountability. This said, it will be argued that if the proposed toolkit is adopted, both *ex-ante* and *ex-post* accountability will be improved.

1. *Conceptualising Automated Decision-Making*

A. Terminology

Automated decision-making refers to a process which is, partly or wholly, automated by a computer system, with the aim of increased efficiency and volume of decision-making.⁴ The term algorithm generally has a wider meaning, constituting a logical process in which a series of rules or conditions are followed to meet defined ends.⁵ Essentially, if conditions A, B and C are met, X will occur as a result. Since all ADM systems are founded on algorithms, the terms in this article are used interchangeably.

Artificial Intelligence takes algorithmic processes a step further, using technological automation to deploy them either somewhat or entirely automatically. The tasks completed by AI can be more complex than more analogue algorithmic processes, which previously depended on human judgment.⁶ A *machine learning algorithm* (MLA) takes this sophistication and complexity a step further still, identifying patterns in vast data sets, rendering them into a manner usable for other purposes such as policy delivery.⁷ While most ADM systems remain as relatively typical algorithms, some constitute incredibly sophisticated MLAs. Such systems have been—and are increasingly being—deployed in service of the aims of the public sector; this may range from the automation of simple tasks (a basic example might be the

⁴ Michael Veale and Irina Brass, ‘Administration by Algorithm? Public Management Meets Public Sector Machine Learning’ in Karen Yeung and Martin Lodge (eds) *Algorithmic Regulation* (OUP 2019) 123.

⁵ Adapted from Sales (n 3) 2.

⁶ *ibid* 4.

⁷ Veale and Brass (n 4) 121.

development of e-filing systems to save time and complexities of paper filing), to extremely complex decision-making structures, which can entirely remove the need for human intervention.

B. Regulatory Challenges

There are undoubtedly benefits to the employment of ADM systems, not least their use in high-volume areas of government; these can boost government capacity and improve service delivery, for instance in fraud detection, police resource allocation and the risk-scoring of prisoners.⁸ However, they are not without their challenges, four of which will be discussed: these are I. legality; II. opacity; III. bias; and IV. contractor accountability. Overcoming these challenges is the central purpose of ADM system regulation.

I. Legality

It is essential that any public sector algorithm accurately represents the underlying legal framework; however, when we go beyond ‘highly specific... semantically un-troubling domains’,⁹ to complex administrative service delivery, accurately representing entire statutory schemes can be challenging. This is particularly so given the value-laden nature of law: concepts like fairness and justice surely cannot be reduced to a few lines of code; they are not quantifiable concepts, they represent value judgments which (human) judges often struggle to grapple with, and can give rise to divergent interpretations.¹⁰ Nor would it necessarily be a good

⁸ *ibid* 137.

⁹ Lilian Edwards and Michael Veale, ‘Slave to the Algorithm? Why a “Right to an Explanation” is Probably Not the Remedy You Are Looking For’ (2017) 16(1) *Duke Law and Technology Review* 18, 24.

¹⁰ As Brad Smith and Carol Ann Browne have written, ‘[h]ow can the world converge on a singular approach to ethics for computers when it

thing to algorithmise such concepts: one of the benefits of such concepts is the ability to deploy them with a degree of flexibility, to meet the justice of the case. The idea of quantifying and weighting fundamentally incommensurable factors, while something that lawyers may be accustomed to, is anathema to most coders.¹¹ Further, an ADM system must continue to appropriately perform its designated functions at the operation stage. A particular risk is that once deployed, an ADM system starts to change the way in which a decision is carried out (e.g. as a result of machine learning), in a manner imperceptible to the public body or the decision subject.¹² To avoid this, it is important that human agency is maintained over the MLA to some degree and that independent (human) judgment of policymakers, who have experience making such decisions, is not undermined.¹³ While it is true that in ADM system design, there will be human involvement to ensure that the system adheres to the statutory and policy framework (since at present, most ADM systems depend on human coders to set their parameters), ensuring that this continues once the code has been written, and the system has been deployed, remains crucial.

cannot agree on philosophical issues for people?', *Tools and Weapons: The Promise and the Peril of the Digital Age* (Penguin 2019) 207.

¹¹ Veale and Brass (n 4) 133. It is also important to note that in addition to coders' inability to measure incommensurable factors, computers themselves may not be able to (at present) replicate the flexibility that humans have in weighing up such factors.

¹² Marion Oswald, 'Algorithm-Assisted Decision-Making in the Public Sector: Framing the Issues Using Administrative Law Rules Governing Discretionary Power' (2018) 376 *Philosophical Transactions of the Royal Society A* 20170359, 14.

¹³ Karen Yeung, 'Algorithmic Regulation: A Critical Interrogation' (2018) *Regulation & Governance* 505, 516.

II. Opacity

The legitimacy of ADM systems risks being undermined given their lack of transparency; such opacity can be intentional or inadvertent. Intentional opacity may result from national security or public policy considerations, and a consequent refusal to disclose the relevant code. For example, it could be dangerous to allow potential terrorists to dissect an ADM system which is used to identify potential flight risks, or to allow potential tax cheats understand the process by which tax records will be selected for audit.¹⁴ Intentional opacity may also arise through an unwillingness of private firms to disclose the system code. For instance, Apple refused to assist the FBI to gain entry into the iPhone of a convicted terrorist, for fear of the chilling effect it may have on private sector innovation.¹⁵

More challenging still is inadvertent opacity. If millions of variables affect a decision, it may be unclear how any single variable ultimately affects the system. In an ADM system deployed by Durham Constabulary, which supports custody officer decision-making, there are 4.2 million decision-points, each dependent on previous data.¹⁶ Hence, even if the inputs

¹⁴ Adapted from Joshua A Kroll, Joanna Huey, Solon Barocas, Edward W Felten, Joel R Reidenberg, David G Robinson and Harlan Yu, 'Accountable Algorithms' (2017) 165 Pennsylvania LR 633, 638-639. See also Veale and Brass (n 4) 138.

¹⁵ Arjun Kharpal, 'Order to hack iPhone for FBI "chilling": Tim Cook' (CNBC, 17 February 2016) <<https://www.cnbc.com/2016/02/17/apple-order-to-hack-iphone-for-fbi-in-san-bernardino-case-chilling-tim-cook.html>> accessed 20 February 2020. See also Apple, 'A Message to Our Customers' (16 February 2016) <<https://www.apple.com/customer-letter/>> accessed 11 March 2021.

¹⁶ Marion Oswald, Jamie Grace, Sheena Urwin and Geoffrey Barnes, 'Algorithmic risk assessment policing models: lessons from the

behind ADM systems are easily understandable, given the ‘black-box’ nature of the processes and the contextual application of the algorithms, outputs are not always predictable.¹⁷ Given that many ADM systems ‘learn’ from contextual data and change their application accordingly,¹⁸ mere inputs will never be sufficient to understand the behaviour of the ADM system; comprehensive algorithmic audit, taking into account source data *and* subsequent contextual data is important. This is difficult to achieve prior to deployment.¹⁹

Therefore, the solution to the problem of opacity cannot simply be transparency through disclosure of the source code. Disclosure will not assist understanding of algorithms, which can be notoriously complex, particularly when we remove the source code from the contextual framework in which the system operates.²⁰ This is as much the case for administrators themselves, as much as it is for members of the public; the inherent complexity of digital governance, and the fact that most civil servants are not programming experts, mean that they will be unlikely to appreciate fully the tools they are using, and the impact that the ADM systems will have. This is not like analogue policy design and implementation, which one would expect to be more intuitively understandable than complex ADM system design. Hence, regulation of ADM systems is important not only to

Durham HART model and “Experimental” proportionality’ (2018) 27(2) *Information & Communications Technology Law* 223.

¹⁷ Marijn Janssen and George Kuk, ‘The challenges and limits of big data algorithms in technocratic governance’ (2016) 33 *Government Information Quarterly* 371, 374.

¹⁸ See discussion of MLA in the first part of this paper.

¹⁹ However, the risks can be mitigated by way of adoption of the methods proposed in the third section of this paper.

²⁰ Anton Vedder and Laurens Naudts, ‘Accountability for the use of algorithms in a big data environment’ (2017) *International Review of Law, Computers & Technology* 206, 215.

ensure the accountability of public sector decision-making to the public, but also to ensure that the ADM systems remain understandable and scrutable tools to *serve*, rather than potentially undermine, policy delivery. As shall be explained in the third section of this paper, technical tools in particular would greatly assist such an understanding of ADM systems.

Even if algorithms are published, there is a risk of their purpose being undermined. This is possible in two senses, as Veale and Brass outline: first, it may permit the targets of ADM systems—such as possible terrorists, in the context of terror risk detection system—to interrogate the methods and datapoints used, so as to ‘play the system’ and escape scrutiny.²¹ If ADM systems are indeed exploited by the very people they are intended to regulate, the effect any risks which do materialise could be catastrophic. This is not to suggest that *human* decision-making processes are not themselves flawed and subject to manipulation; it is just to emphasise the inherent risks of publication of the algorithmic code, which can go beyond mere unintelligibility.

Second, publication of an ADM system’s datapoints may undermine public confidence in the system, particularly if there is a coded-in risk of false negative results.²² This second point is worth underlining; even if we accept the capacity of humans to make mistakes, ‘it seems somehow less palatable that a bureaucracy has decided to deploy a system with a [particular]... rate of false negatives.’²³ While such a rate of false negatives may well be an improvement over the (often flawed and opaque) decision-making processes of humans, and bring consequent benefits to stakeholders, it is the appearance of *intended* failure that the public may find particularly galling. Even with strong

²¹ Veale and Brass (n 4) 138.

²² *ibid.*

²³ *ibid.*

contextual information demonstrating an improvement in decision-making capacity, trust in public bodies to do the right thing may decline.²⁴

III. Bias

A third problem is that of machine bias. In administrative law, a system will show apparent bias, if ‘the fair-minded and informed observer, having considered the facts, would conclude that there was a real possibility of bias’.²⁵ Hence, an ADM system could be biased if it has an internal model which does not produce fair and consistent outputs, in the sense that it consistently benefits or disadvantages a given group; in such a case, apparent bias will likely exist.²⁶ Whereas *prima facie*, ADM systems may *benefit* from the absence of active or passive human bias,²⁷ it is well-established that any code may be value-laden, perhaps (un)intentionally reflecting the coder’s biases. One study, by Kleinberg,

²⁴ A more limited measure of disclosure may well be possible, and may go a long way in accounting for the problem of opacity. However, such disclosure would still have to contend with practical questions of public policy and national security; suitable safeguards such as those that are recommended in the third part of this paper should be adopted, to mitigate such risks.

²⁵ *Porter v Magill* [2001] UKHL 67, [2002] 2 AC 357 [103] (Lord Hope). The more complex the issue, the more ‘informed’ that the fair-minded observer will have to be; this will often result in the test being no more than the actual view of the court being reached: *Hart v Relentless Records* [2002] EWHC 1984, [2003] FSR 36. Further discussion of the test is beyond the scope of this paper; for further information, see Lord Woolf, Jeffrey Jowell, Andrew Le Sueur, Ivan Hare and Catherine Donnelly, *De Smith’s Judicial Review* (8th edn, Sweet & Maxwell 2020) 10-012 ff.

²⁶ Jennifer Cobbe, ‘Administrative Law and the Machines of Government: Judicial Review of Automated Public-Sector Decision-Making’ (2019) 39 *Legal Studies* 636, 654.

²⁷ Veale and Brass (n 4) 126.

Mullainathan and Raghavan, indicated that elimination of bias may never be possible.²⁸ Further, any mistake in the coding, or any widespread social bias reflected in training data used to set up MLAs, may result in biased or discriminatory decisions, which may be difficult to identify given the aforementioned impenetrability of algorithms.²⁹ This risk of entrenching bias is highlighted by a recent threat of judicial review of a Home Office ADM tool which would filter visa applications; the applicants' argument, that the ADM system discriminated on the basis of race, was enough to persuade the Home Office to 'discontinue' and 'redesign' the algorithm. In redesigning the algorithm, the Home Office will consider in particular 'issues around unconscious bias and the use of nationality generally in the streaming tool.'³⁰

IV. Contractor Accountability

Fourth, the development of ADM systems has involved wide-ranging public contracting and involvement of the private sector; this is effectively a continuation of the use of private contractors

²⁸ Jon Kleinberg, Sendhil Mullainathan and Manish Raghavan, 'Inherent Trade-Offs in the Fair Determination of Risk Scores' (2016) <<https://arxiv.org/abs/1609.05807>> accessed 27 December 2020 (cited in Cobbe (n 26) 654).

²⁹ Andrew Tutt, 'An FDA for Algorithms' (2017) 69(1) Administrative LR 83, 105.

³⁰ Henry McDonald, 'Home Office to scrap "racist algorithm" for UK visa applicants' *The Guardian* (London, 4 August 2020) <<https://www.theguardian.com/uk-news/2020/aug/04/home-office-to-scrap-racist-algorithm-for-uk-visa-applicants>> accessed 27 December 2020. For further discussion of the risks of discrimination in ADM systems, see further in this Journal, Gianna Seglias, 'Bias and Discrimination in Opaque Automated Individual Risk Assessment Systems: Challenges for Judicial Review Under the Equality Act 2010' (2021) 10 OUULJ 51.

for the digitisation of the civil service from the 1990s. Recently, for instance, several NHS Trusts have allowed Google's DeepMind to access 1.6 million patient records, in response to Google's willingness to develop health analytic tools for NHS use.³¹ As with any outsourcing contract, it is essential that accountability mechanisms such as qualitative performance targets for contractors and sanctions for non-compliance are included.³² This is particularly important considering the well-documented indifference that private sector bodies often have for administrative law concerns including procedural fairness.³³

2. Regulatory Bodies

In light of the challenges outlined above, regulatory accountability must be obtained prior to deployment. While it will be argued that the optimal method of ADM system regulation is through regulatory tools including data protection laws, impact assessments and codes of practice (Part C), this section outlines

³¹ Madhumita Murgia, 'NHS trusts sign first deals with Google' (*Financial Times*, 19 September 2019) <<https://www.ft.com/content/641e0d84-da21-11e9-8f9b-77216ebe1f17>> accessed 27 December 2020.

³² Regarding possible accountability mechanisms, it is possible to draw analogues with examples given in Pauline Allen, David Hughes, Peter Vincent-Jones, et al., 'Public Contracts as Accountability Mechanisms: Assuring Quality in public health care in England and Wales' (2016) 18(1) *Public Management Review* 20.

³³ UN Special Rapporteur on extreme poverty and human rights, *Digital technology, social protection and human rights* (Report A/74/493, 2019).

the need for regulatory bodies to achieve regulatory accountability.

A. An Algorithm Commission?

Lord Sales has argued that since government lacks the required technical expertise, an expert ‘Algorithm Commission’ should be established, which would scrutinise algorithms before deployment, in a manner similar to pre-legislative scrutiny by the Joint Committee on Human Rights.³⁴ This, he argues, would protect individuals effectively against the risks posed by algorithms.³⁵ His regulator would operate in a manner similar to the utilities regulators, with a key role in protecting consumers and decision-subjects.³⁶ He also suggests that it may operate ‘as a sort of constitutional court’ with the capacity to determine compliance of ADM systems with legal standards.³⁷

Such a suggestion, while novel, is not the first of its kind: Andrew Tutt has argued for an American algorithm regulator similar to the Federal Drug Administration, which would balance technological innovation and public safety.³⁸ This regulator, Tutt suggests, would prevent algorithms from being lawfully deployed until safety testing is completed and disclosure requirements are met. These, he argues, would be adequate safeguards against concerns such as those outlined above. As a form of risk-based regulation, Tutt’s model would categorise algorithms based on their complexity, with corresponding levels of scrutiny afforded to them.³⁹

³⁴ Sales (n 3) 11-12.

³⁵ *ibid.*

³⁶ *ibid* 14.

³⁷ *ibid.*

³⁸ Tutt (n 29).

³⁹ *ibid* 107.

Certainly, the proposals of Sales and Tutt are important. Significantly, from an institutional perspective, a single, centralised regulator is preferable to establishing myriad individual sectoral regulators, whose competences risk overlap and disjointed approaches to ADM systems; far better to have a unitary, comprehensive approach across government. Centralising expertise also ensures that the public sector retains agency over algorithmic deployment, rather than leaving it all to private sector coders.⁴⁰ If this is the case, it would follow that ADM systems are ‘more likely to be designed and implemented in lawful, policy-compliant, and accountable ways.’⁴¹ Additionally, establishing such a regulator would likely bolster individual rights, similar to the Information Commissioner’s role in upholding individuals’ information rights.⁴² In this sense, a regulatory body would be helpful in combatting many of the problems identified in the previous section, particularly with regards to legality and contractor accountability.

However, there are three principal concerns related to Lord Sales’ particular proposal. First, by suggesting that it could constitute an advisor to the courts, a parliamentary aid, *and* an executive agency, this Commission would cut across the separation of powers. No other body (at least since 2005⁴³) has the power to scrutinise government activity and advise on legislative instruments, in addition to enjoying a symbiotic relationship with the courts. Such a power would be tantamount

⁴⁰ Freeman Engstrom and others (n 2) 71.

⁴¹ Sales (n 3) 7.

⁴² Information Commissioner’s Office, Information Commissioner’s Annual Report and Financial Statements 2017-18 (HMSO 2018) 11.

⁴³ The most notable change in the separation of powers was the Lord Chancellor losing their judicial role, see especially Constitutional Reform Act 2005, Part 2 (‘Arrangements to modify the office of Lord Chancellor’).

to declaring the Commission to be judge, jury and (program) executioner! Given the value-laden nature of ADM decisions, it is doubtful whether an ADM regulator ought to have such wide-ranging powers. It is unlikely that the public would place confidence in an organisation which formally belongs in the executive, yet materially impacts the decisions of the other branches of the state. Such doubts were expressed towards tribunals when they remained part of the Whitehall machine; removing them from the executive altogether was required to allay concerns.⁴⁴ Even if we are able to deal with the potential biased or discriminatory application of ADM systems themselves (which has been doubted in the previous section), the appearance of bias in ADM system adjudication is enough to give cause for concern in itself. A more limited regulatory body must be deployed, one which *either* adjudicates *or* administrates.

Second, given the UK's constitutional framework, it is unclear how significant a role the Commission could have in relation to devolution matters: would it have the power, for instance, to reject the implementation of a Scottish ADM system in healthcare, health being a devolved competence? If a devolution matter came before the Supreme Court on the issue, under Lord Sales' approach, we may have the same Algorithm Commission, whose decision is in dispute, advising the Court on the interpretation which ought to be given to the ADM system. This cross-cutting role of the Commission surely cannot be supported. It would be far better to have an ADM system regulator to deal with *narrow* legal questions, within the ADM systems themselves and regarding their application, rather than in the broader constitutional framework in which ADM systems are

⁴⁴ Sir Andrew Leggatt, *Report of the Review of Tribunals* (HMSO 2001).

deployed. The institutional expertise of bodies such as the Supreme Court must be respected.

Third, it is unclear whether a regulator could predict how ADM systems will precisely operate at the *ex-ante* stage. By receiving the source code of an algorithm prior to deployment, the Commission could not properly understand the context in which the data would be applied, or which datapoints are more important to the system. Hence, static review of the source code at the *ex-ante* stage is unilluminating;⁴⁵ it is far from certain that, just by reading the relevant code, regulators would be able to detect any structural issues, such as the misrepresentation or faulty application of the statutory or policy framework, in an ADM system at this stage.⁴⁶ Hence, problems of opacity, as outlined in the previous section, would persist. Far more important would be to identify potential problems before they have the possibility to arise once deployed.

B. A More Limited Regulator

The foregoing analysis is not to suggest that *any* regulator of ADM systems would be so constitutionally problematic. Indeed, a regulatory body ought to be established to ensure the effective deployment of the tools outlined in the next section; every toolkit needs its handyman. It is agreed with Lord Sales that it is sensible to situate the ADM system regulator within the executive, particularly since the executive (rather than, for instance the courts) has the institutional heft which makes it best able to ensure and monitor appropriate policy deployment. Much as the Law Commission is a statutory independent body which is formally part of the executive,⁴⁷ an algorithm regulator would be

⁴⁵ Kroll and others (n 14) 647.

⁴⁶ *ibid* 651.

⁴⁷ See Law Commissions Act 1965.

best suited here, rather than as part of the courts structure. Likewise, it is clear that given the holistic changes driven by AI across government, having several regulators, like the sector-specific ‘Ofdogs’,⁴⁸ would not be particularly effective. The challenges posed by ADM systems are not sector-specific, and do not respect the neat boundaries of government: establishing a single regulatory body, is logical. Where this paper diverges from Lord Sales is regarding the powers and competencies of the regulator. To ensure that the benefits of modern computing may be harnessed by the administrative state, and the risks of such deployment⁴⁹ mitigated, a more restricted regulatory agency—provided it is equipped with the right tools—will be the most effective, and least constitutionally troubling, way of regulating ADM systems.

Rather than forming a new government department which, at least for the present, would overstate the importance of ADM systems, a better method would be to follow existing precedent with regards to executive non-departmental public bodies (NDPBs). These provide important examples from which an ADM-specific regulator, charged with deploying the technical toolkit outlined in Part C, may be developed. Two principal case studies can be highlighted. First, the Equalities and Human Rights Commission (EHRC) is empowered under the Equality Act 2006 to promote and enforce equality and non-discrimination laws across Great Britain. With its wide remit, the EHRC has the power to, *inter alia*: provide advice⁵⁰ and codes of practice⁵¹ to individuals, organisations and public sector bodies; conduct

⁴⁸ Terminology of Carol Harlow and Richard Rawlings, *Law and Administration* (3rd edn, CUP 2009) 67.

⁴⁹ As outlined in the first section of this paper.

⁵⁰ Equality Act 2006, s 13.

⁵¹ *ibid* s 14.

investigations into whether unlawful acts have been committed;⁵² and institute or intervene in relevant legal proceedings.⁵³ Many of the Commission's powers, particularly issuing codes of practice and intervening in legal proceedings, can be read across as effective ways to regulate ADM system deployment by government. Particularly in relation to the problem of bias outlined above, an agency modelled on the EHRC could issue guidance to militate against bias and discrimination in the deployment of ADM systems.⁵⁴ With powers to issue guidance and advice, the regulator could ensure that legal standards are maintained. At the same time, the reserved power to issue or intervene in legal proceedings as an interested party facilitates *ex-post* accountability: technical tools cannot achieve this on their own. Formally part of the executive, albeit institutionally separate (with for instance a ringfenced budget and autonomy from ministerial control), an ADM system regulator modelled on EHRC lines would be unlikely to cut across the separation of powers: it would be able to intervene in cases concerning ADM systems, advising courts on complex issues when invited, but would not become the ultimate arbiter of ADM system disputes in the symbiotic manner that Lord Sales appears to advocate. This would serve to avoid the appearance of bias which must be avoided in the scrutiny of any government policy, let alone in the scrutiny of inherently opaque algorithmic systems.

The second NDPB from which we can learn is the Information Commissioner's Office (ICO), an NDPB responsible for upholding information rights in the public

⁵² *ibid* s 20.

⁵³ *ibid* s 30.

⁵⁴ The deployment of relevant guidance *ex-ante* may well have ensured that the Home Office ADM system used to filter visa applications (above, text at n 30) may well have avoided having discriminatory effect in the first place.

interest.⁵⁵ Its most significant role for present purposes involves supervising adherence to data protection laws;⁵⁶ it may investigate data breaches and pursue enforcement measures against breaching organisations. Its powers do not, however, stop with ‘hard law’ enforcement: the ICO may also audit data controllers, to assess whether they are complying with good practice guidelines.⁵⁷ Given the natural intersection between information rights and algorithmic decision-making, the ICO’s role can inform substantively the formation of an ADM system regulator. Indeed, if we build the ADM regulatory landscape upon existing technologies like the Data Protection Act (DPA) 2018, using the ICO as a touchstone is sensible, and would serve to further the continuing legality and transparency of ADM systems, within the existing regulatory framework.

Using these examples of executive NDPBs with cross-cutting enforcement and advisory powers, it is likely that a similarly constituted algorithm regulator, established along with the proposed regulatory toolkit, will be effective.

3. A New Regulatory Toolkit

Buttressed by the regulator mooted above, it is through the targeted deployment of legal tools that a more sophisticated regulatory landscape for ADM systems can be engineered.

⁵⁵ Information Commissioner’s Office and Department for Digital, Culture, Media and Sport, *Management Agreement 2018-2021* (ICO 2018).

⁵⁶ Data Protection Act 2018, s 115.

⁵⁷ *ibid* s 129.

A. Existing Legislation: GDPR and DPA

Existing legislation provides an effective foundation upon which we can construct the new regulatory landscape. The General Data Protection Regulation (GDPR)⁵⁸ governs the use of personal data. Two important provisions are Article 22, which requires safeguards when fully automatic and significant decisions are made about persons, and Article 28, which mandates certain contractual terms between data controllers and processors. As a result, contracted-out ADM systems do not escape scrutiny. The GDPR was implemented in UK law by the DPA 2018, which includes the provision that where decisions are made solely with automated processing, the data controller must ‘notify the data subject in writing’, who may then ‘request the controller to—(i) reconsider the decision, or (ii) take a new decision that is not based solely on automated processing.’⁵⁹ Hence, where there are concerns around ADM systems, non-automated alternatives may be found.

By giving individuals alternative means of having decisions made about them, these provisions can be applied to suitably cater for transparency and procedural fairness concerns, thus also militating against the problem of opacity which was outlined earlier: if a decision can be made by multiple pathways guided by the same statutory framework, decision-making processes will become more scrutable. They are, however, limited in scope to *personal* data of *particular* individuals, rather than

⁵⁸ Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) [2016] OJ L119/1.

⁵⁹ DPA 2018, s 14(4).

decision-subjects as a class.⁶⁰ A sensible and achievable suggestion could be to ‘bring outsourced ADM which does not involve personal data within its remit’ by extending the GDPR framework.⁶¹ By extending the notification and contract terms requirements discussed above, improved data processing and accountability standards would permeate ADM application, further improving the transparency of such systems.

French law serves as a useful template for development here: under the Administrative Code, individuals have a right to an explanation for ADM-processed decisions made about them. In the explanation, they are provided information about the degree to which the decision was automatic, the data processed, and the particular treatment parameters applied to the situation.⁶² An approach modelled on the existing data protection framework would ensure that systems are designed with accountability at the centre. This would militate against problems of opacity and legality: the failure of ADM system designers to abide by the GDPR’s transparency requirements can be severe.⁶³

A further step could be to require user consent for all automated decision-making. However, in order for this to be effective, users would first need to be able to *understand* the system in order to be able to consent to its presence; this is unlikely, given the innate complexity of AI. For instance, in an ADM system governing the allocation of welfare benefits, to be able to

⁶⁰ Surveillance Studies Network, *Report on the Surveillance Society* (ICO 2006).

⁶¹ Cobbe (n 26) 649.

⁶² Digital Republic Act 2016 (France), ‘Loi n 2016-1321 pour une République numérique’.

⁶³ Under the GDPR, the maximum fine for most breaches is the higher figure of either €10 million, or in the case of an ‘undertaking’, two per cent of its. More severe breaches are liable for twice this: GDPR, Art 83(4)-(5).

understand the hundreds, if not thousands, of relevant variables used by the system to decide to whom to allocate resources; the user would then also need to understand the weight given to particular criteria, and perhaps how their data deviated from the average data of other system users. Given both that ‘the digital citizen has little time for data governance’,⁶⁴ and that attempting to understand mere *simple* algorithms, let alone complex ADM systems with thousands of variables, is incredibly challenging, providing notification to users that they have had a decision made about them through use of an ADM system is more proportionate than requiring consent in every instance. If notified that a decision about them was made electronically, concerned individuals could then request a non-digital alternative decision-making process.⁶⁵ Even where an individual does not understand the nature of the decision-making process, providing them with a decision notice and the right to review (as already occurs in welfare benefits applications⁶⁶) is a proportionate step to ensure administrative accountability.

B. Algorithmic Impact Assessments

The European Parliament has recommended that ‘algorithms in decision-making systems should not be deployed without a prior

⁶⁴ Jonathan A Obar, ‘Big Data and The Phantom Public: Walter Lippmann and the Fallacy of Data Privacy Self-management’ (2015) 2 *Big Data & Society* 1, 13.

⁶⁵ Suggested by the UN Special Rapporteur on extreme poverty and human rights (n 33) 13.

⁶⁶ Claimants have a right to mandatory reconsideration of benefits decisions made about them, see GOV.UK, ‘Challenge a benefit decision (mandatory reconsideration)’ (2020) <<https://www.gov.uk/mandatory-reconsideration>> accessed 28 December 2020.

algorithmic impact assessment'.⁶⁷ Conducting such an assessment, in a similar manner to the GDPR's data protection impact assessment,⁶⁸ would ensure that the most dangerous ADM systems are held back from release, or are subjected to heightened operational scrutiny. Although the operational risk of many ADM systems cannot be comprehensively tested or modelled prior to its application, programmers can raise red flags, which decision-makers can then effectively monitor, so that identified risks do not materialise. In the event that they do materialise, the impact assessment will have established system-specific procedures to combat their effects. Of course, being able to identify legal risks beyond those inherent in all algorithms requires a degree of legal literacy; this is why it is important for lawmakers, policymakers and technical specialists to learn from one another and understand each other's domains.⁶⁹ Directly accounting for the risks outlined in Part A, particularly bias and legality, impact assessments are an essential part of any contemporary regulatory toolkit. As a core tool of *ex-ante* regulation, by militating against risks from emerging in the first place, the *ex-post* accountability burden placed on any regulatory agency would necessarily be less intensive than if there was no such risk-identification process.

C. Codes of Practice

The adoption of codes of practice by public sector bodies would improve the consistency of decision-making and likely improve transparency: such codes are often regarded as a crystallisation of

⁶⁷ European Parliament resolution of 12 February 2019 on a comprehensive European industrial policy on artificial intelligence and robotics (2018/2088(INI)) [2020] OJ C449/37 [154].

⁶⁸ GDPR, Art 35(7).

⁶⁹ Richard Susskind, *Tomorrow's Lawyers: An Introduction to Your Future* (OUP 2017) 184-190.

good governance norms.⁷⁰ The UK Government adopted a Data Ethics Framework in 2018, which is intended to guide public sector ‘policy and service design’.⁷¹ Reading this across to ADM systems, it would provide for a proportionate use of data in designing systems, safeguarding that decisions are reached using only relevant considerations. By requiring that these good governance norms are imposed as contractual terms on the contractors who design ADM systems, codes of practice could have normative force and help ensure contractor accountability. This could be further guaranteed by public bodies adopting the recommendations of the House of Commons Science and Technology Committee, that ‘Government should produce... a list of where algorithms with significant impacts are being used’, and make it available for external analysis.⁷² Such a voluntary policy of disclosure is likely to strengthen public confidence in ADM systems, particularly if the list becomes available for scrutiny by third-sector organisations.

One particular duty which could be included in the codes of practice, as proposed by Lord Sales, is a duty to ‘have regard’ for the interests of decision subjects.⁷³ Such a wide-ranging duty is desirable, ensuring a wider range of considerations are taken into account by algorithms. Further, it reflects the legislative zeitgeist: the Well-being of Future Generations (Wales) Act 2015 establishes a similar ‘take account’ duty to all administrative action

⁷⁰ Cary Coglianese and David Lehr, ‘Regulating by Robot: Administrative Decision-Making in the Machine-Learning Era’ (2017) 105 *Georgetown LR* 1147, 1211.

⁷¹ Department for Digital, Culture, Media and Sport, ‘Data Ethics Framework’ (HMSO 2018) <<https://www.gov.uk/government/publications/data-ethics-framework>> accessed 20 February 2020.

⁷² *Algorithms in Decision-Making* (HC 351, 2018).

⁷³ Sales (n 3) 10.

of Welsh public bodies.⁷⁴ Looking across to other codes of practice, such as the EHRC guidance on the Equality Act 2010,⁷⁵ we see that the use of example scenarios in the guidelines can also be illuminating and ensure decision-makers act according to best practice.

Although codes of practice are non-binding, this is not necessarily a drawback to their efficacy. Compliance with codes of practice is likely to improve public trust in ADM systems and, by incorporating such codes into public contracts for ADM system design, further encourage good faith compliance. It is not the case that hard law enforcement will always be the most effective way of ensuring accountability: when we consider codes of practice alongside existing legislative tools and impact assessments, a hybrid ‘carrot and stick’ approach is more likely to ensure compliance. Even if it were only government *policy* to ensure that certain good governance norms are required in all public contracts for the design of ADM systems, this would go a long way in improving contractor accountability. Given that the ultimate aim of this toolkit is to improve *ex-ante* accountability, using codes of practice to crystallise norms of good algorithmic governance—both within Whitehall and among its private sector system designers—is likely to result in better-coded ADM systems. This, in turn, will mean that less acute *ex-post* enforcement will be needed.

D. Technical Tools

Fourth, there are particular technical tools available to regulators which can verify automated decisions and ensure compliance with

⁷⁴ See especially s 5.

⁷⁵ EHRC, ‘Equality Act codes of practice’ (2019) <<https://www.equalityhumanrights.com/en/advice-and-guidance/equality-act-codes-practice>> accessed 27 April 2020.

legal standards. Designing ADM systems with external scrutiny in mind must do more than merely examining the inputs and outputs of the system; the ADM system itself must be scrutable. This is particularly important, given the necessary limitations in *ex-ante* scrutiny, and the difficulty in understanding individual decisions unless we understand the ADM system as a whole. Three technical tools in particular are suggested,⁷⁶ each of which will serve to promote algorithmic accountability. First, software verification tools allow coders and auditors to test whether an ADM system has certain properties. For instance, in an ADM system designed to implement a statutory scheme, it may be tested using software verification tools whether the system has taken into account all relevant considerations as enumerated in the legislation. Coders will also be able to reason how the system will behave under all conditions:⁷⁷ if one of the inputs into the algorithm changes, it would remain possible to verify how it would affect its outputs. Methods of software verification can vary, from exhaustive *ex-ante* testing for all possible variables, to designing a complementary algorithm which will decode ADM systems to prove that the specified properties exist.⁷⁸ Testing for particular properties is of particular importance when decision subjects are categorised according to those properties, particularly if these properties are protected characteristics which public bodies must protect.⁷⁹ Through this method, we could determine, for instance, whether a given algorithm is likely to give rise to a biased or discriminatory application. In testing any new ADM systems similar to its visa-sorting algorithm,⁸⁰ the Home Office could deploy such tests to ensure that it functions in a non-

⁷⁶ Drawn from Kroll and others (n 14).

⁷⁷ *ibid* 662.

⁷⁸ *ibid* 663.

⁷⁹ Equality Act 2010, Chapters 1 and 2.

⁸⁰ See text at n 30.

discriminatory manner. By doing so, Whitehall would continue to make efficiency gains, while having tested comprehensively that the relevant properties are present in its ADM systems.

Second, cryptographic commitments will allow auditors to prove *ex-post* that a given program has performed in accordance with the original underlying commitments, staying true to its legal purpose.⁸¹ These are the digital equivalent of drafting a secure document with a list of particular properties that we wish for a given policy to have; the policy could then be checked later on in relation to this document, to ensure that it has remained true to its undertakings. Whereas parliamentary committees scrutinise the acts of departments, complementary technical tools could be used to test for the legality of ADM systems once they are deployed (and could themselves be used as evidence to parliamentary committees, in their scrutiny of departmental activity). Secure documents would normally be held off-site and checked by an auditor; in the present case this could be facilitated by having the algorithmic regulatory body periodically test ADM systems in relation to these cryptographic commitments. The more detailed the cryptographic commitments that are made in the first place, the more that must be done to ensure fidelity to them as an ADM system operates. This counters the fears that an ADM system may no longer represent the legal framework underpinning the system, thus ensuring its deployment remains lawful, while enabling the source code to remain undisclosed. This further ensures that the software used to implement the policy is determined and recorded prior to the implementation stage, meaning that neither the policy nor its application are affected by external factors.⁸² Such tests would remain important whenever an ADM system is deployed in a sensitive area,

⁸¹ Kroll and others (n 14) 666.

⁸² *ibid* 667.

including crime detection and prevention,⁸³ and the identification of flight risks.⁸⁴

Third, zero-knowledge proofs—a further cryptographic tool—allow the decision-maker to prove that an ADM system has a certain property, without revealing the nature of that property.⁸⁵ This is helpful when we consider that transparency *per se* is not necessarily useful. If a decision-maker makes a trio of commitments—to policy, inputs, and a decision reached—then a zero-knowledge proof will allow the verification that the three elements correspond: that the correct policy was applied to the correct input, to reach the stated outcome.⁸⁶ This will ensure consistent policy delivery, and facilitate internal departmental auditing, to ensure policies stay congruent with their parent department's remit.

As a result of these technical tools, if applied, ADM systems would become more readily testable and auditable, and their decisions would become easier to explain. All of this could be achieved without compromising the functionality of the systems. By incorporating accountability into system design, it will ensure a more assured *ex-ante* scrutiny of algorithms. At the same time, tools such as these would render unnecessary and disproportionate the broader institutional approach argued for by Lord Sales: by ensuring more effective *ex-ante* regulation of ADM systems as a function of their design, the need to develop a broad public sector regulator with wide ranging adjudicative, advisory and rule-making powers, would be unnecessary.

⁸³ The Durham HART model is a good example of such a model which must remain congruent with its original policy underpinnings. For discussion of the model, see Oswald and others (n 16).

⁸⁴ Adapting Veale and Brass (n 4) 138.

⁸⁵ Kroll and others (n 14) 668.

⁸⁶ *ibid.*

E. Impact

In addition to supporting *ex-ante* scrutiny of ADM systems, the four regulatory techniques identified would strengthen *ex-post* scrutiny. Internal audit would become more straightforward if cryptographic commitments we used, as these facilitate intra-system testing; administrative accountability would become more automatic; and possible risks could be identified by codes of practice and prophylactically mitigated by impact assessments. All of this is even more assured if an algorithm regulator is deployed. Of course, this regulator, while responsible for ensuring the effective deployment of the toolkit *ex-ante*, would also be able to review *ex-post* any ASM systems developed, considering the prior risk assessment and cryptographic tools. Like the audit capability of the ICO, this would likely strengthen compliance measures. Finally, with powers to institute judicial review proceedings, the regulator, supported with its toolkit, could bring actions, particularly in relation to the risks of bias and discrimination, which are difficult to control for before any contextual data is fed into the ADM system. The crucial distinction between the proposed regulator here, and that of Lord Sales, is twofold. First, the proposed regulator would necessarily be limited, and would necessarily play second fiddle to the primary regulatory *mechanisms* developed. Second, the adjudicative role which was proposed by Lord Sales would not be deployed: it is not for an administrative agency itself to decide on the intrinsic legality of ADM systems; this is primarily a role for the courts. It would be more appropriate for the regulator to issue the (non-binding) industry codes of practice, and ensure scrutiny of ADM systems *ex-post*; like the EHRC, it may then intervene in legal challenges. This would ensure the regulatory body played an important—but necessarily limited—role in challenges to ADM system adjudication.

Conclusion

In this article, it has been argued that while deploying a regulatory body to ensure accountability for ADM systems may be effective, a superior and more sophisticated method is to apply an expanded regulatory toolkit consisting of existing digital governance rules, algorithmic impact assessments, codes of practice, and cryptographic proof tools. By marrying up existing bodies and procedures with new innovations, accountability techniques may develop to facilitate the beneficial use of ADM systems, rather than inhibiting innovation. Going forward, this framework may serve as a stepping-stone on the path towards comprehensive digital and algorithmic accountability; many of the tools outlined could readily be applied at a local level, and across to the private sector. To avoid the new generation of digital citizens sleepwalking into becoming ‘electric sheep’,⁸⁷ with all major decisions made about our lives concluded by algorithm, regulators must ensure, and citizens must insist on, comprehensive digital accountability. Given the limitless opportunities afforded by AI and ML, well-crafted regulation offers us the ability to safely harness new technologies, while remaining cognisant of their risks.

⁸⁷ Philip K Dick, *Do Androids Dream of Electric Sheep?* (Doubleday 1968).

*Bias and Discrimination in
Opaque Automated Individual
Risk Assessment Systems:
Challenges for Judicial Review
under the Equality Act 2010*

Gianna Seglias*

Abstract—Public authorities’ use of automation to assist decision-making poses a novel challenge for administrative law principles, which were developed with reference to human decision-making processes. This article explores one particular aspect of this challenge. It considers how public authorities utilise automated systems to assess the statistical risk of individuals engaging in particular types of undesirable behaviour, and explains how these technologies may facilitate biased outcomes. It then analyses how judicial review under the Equality Act 2010 could be used to challenge the use of such systems, focussing in particular on sections 29 and 149. Finally, it outlines the evidential challenge posed by algorithmic opacity and the potential of the public sector equality duty to mitigate this difficulty.

* University College London. I am grateful to the OUUIJ editorial team for comments on earlier drafts. All errors remain my own.

Introduction

The combined effect of the public sector's enthusiasm for new technologies¹ and the legacy of New Labour-era governance focussed on strategic risk management² is that public authorities have been eager to adopt algorithmic tools to assess the statistical risk of individuals behaving in undesirable ways. While these technologies have the potential to facilitate efficiently targeted intervention, concerns have been raised about the potential for biased outcomes.³ Where the burden of technological bias falls on a group possessing a protected characteristic under the Equality Act 2010 (EA 2010), decisions made using these tools may constitute unlawful discrimination under UK equality law. This article will consider whether administrative law is sufficiently equipped to protect individuals from discriminatory decisions made with the assistance of machine learning automated individual risk-assessment systems (AIRAS) and argue that the existing grounds of review under the EA 2010 may have significant potency. Part 1 introduces the workings and uses of AIRAS, considers their discriminatory effects, and explains the evidential challenge posed by algorithmic opacity. Part 2 explores the potential of EA 2010 sections 29 and 149 in challenging the use of discriminatory AIRAS, noting in particular the potential of the public sector equality duty (PSED) to create much-needed transparency. Part 3 concludes that although piecemeal review under the EA 2010 may be an effective mechanism for combating

¹ Government Digital Service, *Government Transformation Strategy* (2017) <www.gov.uk/government/publications/government-transformation-strategy-2017-to-2020> accessed 5 April 2021.

² Carol Harlow and Rick Rawlings, *Law and Administration* (3rd edn, CUP 2009) 73ff.

³ For example see Joshua Kroll et al., 'Accountable Algorithms' (2017) 165 *University of Pennsylvania Law Review* 633.

discrimination in AIRAS in the short term, the creation of *ex ante* accountability mechanisms could ultimately offer a more comprehensive and principled response to the questions raised by algorithmic bias.

1. *How AIRAS Discriminate*

Risk assessment systems are a subset of automated decision-making systems which use algorithms to make decisions with reduced or no human input. AIRAS are concerned with providing a decision-maker with information about whether an individual is 'at risk' of having a certain characteristic or carrying out some undesirable activity. These systems calculate this risk by processing known data. For example, if the relevant characteristic or activity, or *output variable*, is child abuse, the model will consider past recorded instances of abuse to determine which other factors, or *input variables*, correlate to the occurrence of this event. Some AIRAS are systems in which the input variables are specified and their statistical weight in contributing to the output is determined by programmers. However, other systems utilise 'machine learning', a form of artificial intelligence through which a system autonomously develops its decision-rules by being run over a set of training data.⁴

⁴ *ibid* 638.

A. Uses of AIRAS

AIRAS are already in use by various UK public authorities.⁵ In the criminal justice system, they are used to make decisions about policing, charging, and parole.⁶ They have also been used by children's services to identify children at risk of abuse and to make safeguarding decisions.⁷ A potential area for expansion of AIRAS is in immigration and border control, where predictive analytics could be used to determine which individuals are likely to become over-stayers or pose a danger to national security. In the future, AIRAS are likely to expand to other areas of citizen-state interaction, particularly where public authorities working with reduced budgets are looking to implement more cost-effective decision-making procedures.

B. Potential for Discriminatory Decisions

Machine learning AIRAS are frequently biased. Under a broad understanding of the term, some form of 'bias' is inherent to such systems because they use data about *groups* of people to predict the behaviour of *individuals*. However, many AIRAS are also 'biased' in a more narrow sense; this latter understanding, which

⁵ For examples see Lina Dencik et al., 'Data Scores as Governance: Investigating uses of citizen scoring in public services' (2018) <<https://datajustice.files.wordpress.com/2018/12/data-scores-as-governance-project-report2.pdf>> accessed 5 April 2021.

⁶ See Michael Veale et al., 'Algorithms in the Criminal Justice System' (Law Society 2019) <www.lawsociety.org.uk/en/topics/research/algorithm-use-in-the-criminal-justice-system-report> accessed 5 April 2021.

⁷ Hackney Council ended a pilot programme, the Early Help Profiling System, in 2019. For an international comparison, see the Allegheny Family Screening Tool, described in Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police and Punish the Poor* (St Martin's Press 2018).

will be adopted in this paper, refers to systems which make predictions that are systematically too high or too low for specific subgroups.⁸ Where the relevant group corresponds to individuals possessing a ‘protected characteristic’ (as defined in sections 4-12 EA 2010), this bias may constitute discrimination in the legal sense. Kroll et al. identify three ways in which automated decision-making systems ‘simultaneously systematise and conceal discrimination’.⁹ Firstly, they may be trained on data sets reflecting past prejudice, meaning that the model treats prejudiced decisions as an example to learn from.¹⁰ Secondly, they can build in discrimination through model construction. This may arise in particular at the stage where data is selected to be considered by the model (feature selection).¹¹ For example, it is common in AIRAS to identify membership of a protected class as an input, because protected characteristics may indeed correlate with the relevant outcome risk variable. In a child abuse prediction model, for instance, people of a particular race or gender may be statistically more likely to have their children taken into foster care following child abuse incidents. Even if these characteristics are not fed into the algorithm, they may ‘sneak back in’ through proxies, which are included variables containing ‘signals’ predicting the excluded variable.¹² For example, postcode may

⁸ Partnership on AI, ‘Report on Algorithmic Risk Assessment Tools in the US Criminal Justice System’ (2019)

<www.partnershiponai.org/report-on-machine-learning-in-risk-assessment-tools-in-the-u-s-criminal-justice-system/> accessed 5 April 2021, 15.

⁹ Kroll et al. (n 3) 680.

¹⁰ Solon Barocas and Andrew D Selbst, ‘Big Data’s Disparate Impact’ (2017) 104 *California Law Review* 671.

¹¹ Kroll et al. (n 3) 681.

¹² Michael Veale and Lilian Edwards, ‘Slave to the Algorithm? Why a “right to an explanation” is probably not the remedy you are looking for’ (2017) 16 *Duke Law & Technology Review* 18, 29.

predict race.¹³ Thirdly, systems may ‘mask’ intentional discrimination by a prejudiced decision-maker or model developer.¹⁴ Even where the final human decision-maker is themselves unbiased, discrimination in the system is replicated if the decision-maker overly relies on the automated decision.¹⁵

C. Case Studies

There are two AIRAS about which there has been significant commentary: COMPAS, in the USA, and HART, in the UK. Both are used in the criminal justice system to assess the risk of recidivism and each has been accused of facilitating biased decisions. Reference will be made to these case studies throughout the paper.

I. COMPAS

The Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) system is used by courts in several American jurisdictions to assess recidivism risk for the purpose of various criminal justice decisions, including sentencing. Among other things, the system provides both a General Recidivism Risk scale, for which the output variable is an arrest within two years of the intake assessment, and a Violent Recidivism scale, for which the output variable is an arrest within

¹³ Black and minority ethnic communities are more concentrated in certain postcodes - see Office for National Statistics, ‘2011 Census’ <www.nomisweb.co.uk/census/2011/ks201ew> accessed 5 April 2021.

¹⁴ Kroll et al. (n 3) 682.

¹⁵ See comments on ‘automation bias’ by Reuben Binns as reported in Alice Irving, ‘Rise of the algorithms’ (2019) (*UK Human Rights Blog*, 4 November 2019) <<https://ukhumanrightsblog.com/2019/11/04/rise-of-the-algorithms/>> accessed 5 April 2021.

two years for an offence on a person.¹⁶ Not much is known about the algorithm employed by COMPAS, as the software was privately developed by Northpointe (now trading as Equivant) and continues to be owned by the company. The risk scores are based on public criminal records¹⁷ and answers to a 137-question survey which tracks both static factors, such as criminal history, and dynamic factors, such as attitudes towards right and wrong.¹⁸

A 2016 investigation by ProPublica on the use of COMPAS in Broward County, Florida, concluded that the system disadvantaged Black defendants.¹⁹ While the tool made errors at roughly the same rate for Black and white defendants, the *direction* of error made differed according to defendants' race. The system was more likely to mislabel Black defendants as high risk and conversely more likely to mislabel white defendants as low risk. Even when accounting for criminal history, age, and gender, Black defendants were 77 percent more likely to receive a higher score on the Violent Recidivism scale and 45 percent more likely to receive a higher score on the General Recidivism Risk scale. This disparity is illustrated in the ProPublica study by comparing two real-life examples of erroneous predictions made in relation to two petty theft suspects: a white defendant who was rated low-risk despite previous convictions for armed robbery and who subsequently committed burglary, and a Black defendant with

¹⁶ Northpointe, 'COMPAS Risk & Need Assessment System: Selected Questions' (2012) <www.northpointeinc.com/files/downloads/FAQ_Document.pdf> accessed 16 June 2020.

¹⁷ *Loomis v Wisconsin* 881 NW2d 749 (Wis 2016) [55].

¹⁸ Northpointe, 'Risk Assessment' (2012) <www.documentcloud.org/documents/2702103-Sample-Risk-Assessment-COMPAS-CORE.html> accessed 5 April 2021.

¹⁹ Julia Angwin et al., 'Machine Bias' (*ProPublica* 23 May 2016) <www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> accessed 5 April 2021.

only a juvenile misdemeanour record who was rated high risk, but has not since been charged with any further offences.

In response to the investigation, Northpointe ‘strongly reject[ed]’ the allegation of bias, maintaining that COMPAS was ‘equally accurate’ for all races,²⁰ since it correctly classified people as recidivists at the same rate.²¹ These contradictory framings of the model—as either unbiased because it makes errors at an equal rate for both groups, or as biased because the burden of false positives falls disproportionately on one group—reflect the ‘impossibility theorem of fairness’. This is a statistical rule which holds that where there are different rates of occurrence of an output variable between groups, a model can either have equal predictive accuracy for both groups *or* achieve equal rates of false positives and negatives for both groups, but cannot simultaneously satisfy both fairness criteria.²² Accordingly, even an instrument seemingly free from predictive bias may result in a ‘disparate impact’ in systems where a higher risk score results in disadvantage for the subject.²³

II. HART

Durham Constabulary’s Harm Assessment Risk Tool (HART) is used at the point of arrest to identify individuals at low or moderate risk of recidivism, who are to be referred to the

²⁰ William Dieterich, Christina Mendoza and Tim Brennan, ‘COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity’ (Northpointe 2016) <http://go.volarisgroup.com/rs/430-MBX-989/images/ProPublica_Commentary_Final_070616.pdf> accessed 21 April 2021, 1.

²¹ *ibid* 7.

²² Alexandra Chouldechova, ‘Fair prediction with disparate impact: A study of bias in recidivism prediction instruments’ (FATML conference, New York, November 2016).

²³ *ibid* 1.

‘Checkpoint’ programme rather than being charged. HART utilises 34 input variables and employs a random forest algorithm (a type of machine learning) to classify offenders into three risk groups. The model intentionally favours ‘cautious’ errors, meaning it is more likely to over-predict risk than it is to under-predict.²⁴ The system was developed in-house by Durham Constabulary and details of the model’s construction are publicly available.²⁵

While there has been no comprehensive investigation of HART similar to that conducted by ProPublica on COMPAS, HART may nevertheless be discriminatory since it utilises information related to the protected characteristics of race, gender and age. Gender and age are used directly as inputs, and two of the input variables include postcode data²⁶ which can act as a proxy for race.²⁷ Since communities of colour in the UK, particularly Black communities, are frequently over-policed,²⁸ use of such data may generate a ‘feedback loop that may perpetuate

²⁴ Marion Oswald et al., ‘Algorithmic risk assessment policing models: lessons from the Durham HART model and “Experimental” proportionality’ (2018) 27 *Information and Communications Technology Law* 223, 228.

²⁵ See Sheena Urwin, ‘Algorithmic Forecasting of Offender Dangerousness for Police Custody Officers: An Assessment of Accuracy for the Durham Constabulary Model’ (MSc thesis, University of Cambridge 2016) <www.crim.cam.ac.uk/system/files/documents/sheena-urwin-thesis-12-12-2016.pdf> accessed 5 April 2021, Appendix B.

²⁶ Oswald et al. (n 24) note that Durham Constabulary intended to remove one of two postcode predictors but it is unclear whether this took place.

²⁷ See the 2011 census (n 13).

²⁸ Home Affairs Committee, *The MacPherson Report – 10 Years On* (HC 2008-2009, 427).

or amplify existing patterns of offending...leading to an ever-deepening cycle of increased police attention'.²⁹

D. System Opacity and Evidential Difficulties

AIRAS may, however, pose evidential difficulties for litigants seeking to establish discrimination, due to algorithmic opacity. Cobbe differentiates between intentional, illiterate, and intrinsic opacity.³⁰

Intentional opacity refers to the situation where the workings of a system are deliberately concealed, either because the technology provider is a for-profit entity or because the public authority wishes to prevent individuals from 'gaming' the system. For example, the algorithm used in COMPAS is a trade secret owned by Equivant, and, following a constitutional challenge to the tool, the Wisconsin Supreme Court noted that without access to the source code it could not evaluate 'how the risk scores are determined or how the factors are weighed'.³¹

Even where the algorithm utilised in an AIRAS is publicly available, it may be understandable only to technological experts (*illiterate* opacity) or not understandable to humans at all (*intrinsic* opacity). Like many machine learning systems, HART is intrinsically opaque due to its complexity. Oswald et al. note that the algorithm is 4.2 million 'nested and conditionally-dependent decision points',³² meaning that any one variable has no direct impact on the result and that discriminatory effects may be so small that they are not noticeable in any particular case, but give

²⁹ Oswald et al. (n 24) 228.

³⁰ Jennifer Cobbe, 'Administrative law and the machines of government: judicial review of automated public-sector decision-making' (2019) 39 *Legal Studies* 636.

³¹ *Loomis v Wisconsin* (n 17) [51].

³² Oswald et al. (n 24) 228.

rise to significant ‘cumulative disadvantage’ across the system.³³ System opacity may be further compounded by data protection obligations. For example, Veale et al. note that the kind of investigation conducted into COMPAS would not be possible in the UK as the training data used would likely not be released under UK freedom of information law.³⁴

The high level of opacity present in AIRAS harms the accountability of public authorities, which may be understood as the obligation of public officials to ‘explain and justify their conduct’ so that these actions may be subjected to scrutiny and, where appropriate, result in consequences.³⁵ Accountability between the governing and the governed is disrupted where decision-making occurs through procedures which the public cannot understand. In particular, legal accountability through judicial review may be hindered where decision-making processes become inscrutable or indecipherable. Because public law review mechanisms are ‘primarily concerned with decision-making processes’,³⁶ algorithmic opacity may mean that claimants are unable to gather the evidence necessary to establish that a decision is affected by public-law error.

The urgency of legal reform to facilitate transparency and explainability of algorithmic systems has been recognised in the literature,³⁷ but this article will not offer a policy exploration of

³³ See Oscar Gandy, ‘Engaging rational discrimination: exploring reasons for placing regulatory constraints on decision support systems’ (2010) 12 *Ethics and Information Technology* 29.

³⁴ Veale et al. (n 6) 21, note 66.

³⁵ Mark Bovens, ‘Analysing and Assessing Accountability: A Conceptual Framework’ (2007) 13 *European Law Journal* 447.

³⁶ Cobbe (n 30) 649.

³⁷ See for example Swee Leng Harris, ‘Data Protection Impact Assessments as rule of law governance mechanisms’ (2019) 2 *Data & Policy* 1.

how to remedy opacity in AIRAS or automated systems more generally. It will concern itself with the narrower question of how EA 2010 sections 29 and 149 might be utilised to challenge discriminatory AIRAS, and will consider the impacts of opacity on the prospect of success of challenges under those provisions. The following section will argue that existing legal principles under the EA 2010 can offer significant protection to claimants, and will conclude that opacity is unlikely to frustrate such claims. Instead, opacity may in some instances work in claimants' favour, by allowing them to take advantage of legal principles to lower their evidential burden while simultaneously rendering it more difficult for public authorities to meet the procedural and substantive burdens placed upon them by equality law. Accordingly, judicial review claims under the EA 2010 may not only offer relief to claimants, but could contribute to the broader aim of combating algorithmic opacity.

3. Legality Review Under the Equality Act 2010

There are, in theory, many avenues of judicial review through which claimants may challenge either individual discriminatory decisions made through AIRAS, or policy decisions on the design and use of such tools. This article, however, will consider only review under the EA 2010, which is directly concerned with equality issues and would accordingly provide litigants with the most straightforward route. While the EA 2010 has largely been utilised for civil claims against the state, a decision which fails to comply with the EA 2010 may also be judicially reviewed as *ultra*

viris.³⁸ In particular, a decision is unlawful where it constitutes unlawful discrimination under section 29(1), (2) or (6),³⁹ or fails to comply with the PSED in section 149(1).

A. Prohibition on Direct and Indirect Discrimination

Decisions may be challenged as violating EA 2010 section 29 where they either directly or indirectly discriminate against the claimant.

I. Substantive Requirements

Section 13(1) defines direct discrimination as less favourable treatment ‘because of’ a protected characteristic as defined in sections 4-12. The Supreme Court held in *R (Coll) v Secretary of State for Justice*⁴⁰ that a greater *risk* of a disadvantage which materialises for the claimant is sufficient to make out direct discrimination even if not all members of the protected group suffer the disadvantage; in that case, it sufficed that women were at a higher risk of being required to live in premises further from their home after release from prison due to the lower number of facilities for women. Allen and Masters argue that following *Coll*, it should be possible to argue that risk asymmetry under an automated system gives rise to less favourable treatment where people possessing a protected characteristic are at higher risk of an unfavourable score.⁴¹ This focus on risk dovetails with the

³⁸ EA 2010 s 113(3)(a).

³⁹ The prohibition against discrimination in the provision of services in s 29(1)-(2) also includes provision of service in the exercise of a public function: s 31(3).

⁴⁰ [2017] UKSC 40.

⁴¹ Robin Allen and Dee Masters, ‘In the Matter of Automated Data Processing in Government Decision Making’ (*AI Law Hub*, 7

public law doctrine of systemic unfairness developed in recent cases which places emphasis on a ‘risk of unfairness, rather than its realisation’.⁴² Courts will not be receptive to arguments that actual statistical difference between populations means there is no discrimination, since equality law mandates that ‘individuals should be treated as individuals, and not assumed to be like other members of a group’.⁴³ For example, since the HART algorithm uses age and sex as input variables, claimants should be able to establish direct discrimination if the system disadvantages them on these grounds (though age discrimination can be objectively justified under section 13(2)).⁴⁴

However, a challenge may be more difficult where a protected characteristic is only considered through a proxy variable, such as the inclusion of race through the postcode variable in HART. Lady Hale reaffirmed in *Coll* that where the criterion used by the discriminating actor is a mere proxy for the protected characteristic, there must be ‘exact correspondence’ between the protected characteristic and the ‘disadvantaged class’

September 2019) <www.cloisters.com/wp-content/uploads/2019/10/Open-opinion-pdf-version-1.pdf> accessed 5 June 2021, 14.

⁴² Abi Adams-Prassl and Jeremias Adams-Prassl, ‘Systemic Unfairness, Access to Justice and Futility: A Framework’ (2020) 40 *Oxford Journal of Legal Studies* 561, 567.

⁴³ *R(E) v Governing Body of JFS* [2009] UKSC 15 [90].

⁴⁴ While sch 3 para 3(1)(c) provides that s 29 does not apply to ‘a decision not to commence or continue criminal proceedings’ the wording of this section implies that *positive* decisions to charge prosecute are not excluded. It should be noted, however, that a system like COMPAS used in sentencing decisions would likely not be reviewable since most judicial decisions are not subject to judicial review and in any event sch 3 para 3(1)(a) excludes judicial functions from consideration under s 29.

(i.e. the group of people disadvantaged by the provision).⁴⁵ It is unlikely that this test would be met in most cases, since proxy variables do not usually track protected characteristics perfectly. As a result, some individuals who do not possess the protected characteristic suffer the disadvantage while some individuals who possess the protected characteristic do not. This may be illustrated by considering HART. Assuming that the tool assigns individuals a less favourable risk score because they reside in a postcode with higher populations of ethnic minority residents, some ethnic minority residents who live in majority-white postcodes will not suffer this disadvantage. Conversely, some white residents of postcodes with higher populations of ethnic minority residents will be disadvantaged even though they do not share the protected characteristic.

In proxy cases, claimants may instead seek a declaration that use of an AIRAS constitutes indirect discrimination contrary to section 19. Following section 19(1), a person or organisation discriminates against an individual indirectly by applying to them a provision, criterion or practice (PCP) which, though apparently neutral, is in practice discriminatory. A PCP is to be construed widely,⁴⁶ and so could include the algorithm itself, the policy used to guide its implementation, or its training dataset.⁴⁷ A claimant would need to demonstrate that the PCP places them and others sharing their protected characteristic at a particular disadvantage when compared to those who do not share the protected

⁴⁵ *Coll* (n 40) [28]-[29].

⁴⁶ Equality and Human Rights Commission, 'Services, Public Functions and Associations: Statutory Code of Practice Equality Act 2010' (2011) <www.equalityhumanrights.com/sites/default/files/servicescode_0.pdf> accessed 5 April 2021, 70.

⁴⁷ *Allen and Masters* (n 41) 16.

characteristic.⁴⁸ Helpfully for challenges to AIRAS, UK courts have held that a relatively small discrepancy is sufficient to show disadvantage⁴⁹ and that disparate impact can be established on the basis of statistical evidence.⁵⁰ If it were shown, for example, that ethnic minority individuals were statistically more likely than white individuals to be given a higher risk score by HART and accordingly less likely to be referred to Checkpoint, a claimant could argue the tool (or related policy guidance) to be a *prima facie* discriminatory PCP.

While the two types of discrimination are distinct from one another and ‘mutually exclusive’,⁵¹ so that a finding of direct discrimination ‘rules out’ a finding of indirect discrimination,⁵² they are frequently pleaded in the alternative. In challenges to AIRAS, which are likely to focus on decision-making *systems* and *risk* of disadvantage evidenced through statistical disparities, the two types of discrimination may be difficult for courts to distinguish in practice. As noted above, they are likely to find the dividing line between the two wrongs to be whether the protected characteristic is an input variable or merely considered through a proxy.

The distinction between direct and indirect discrimination is of practical importance, since indirect discrimination can be justified where the PCP is shown to be a proportionate means of achieving a legitimate aim.⁵³ Helpfully to claimants, it will be difficult for the public authority to justify a

⁴⁸ s 19(2).

⁴⁹ *London Underground v Edwards (No 2)* [1999] ICR 494 (CA).

⁵⁰ *Essop and ors v Home Office* [2017] UKSC 27 [28].

⁵¹ Bob Hepple, *Equality: The Legal Framework* (2nd edn, Hart 2014) 81.

⁵² *Coll* (n 40) [43].

⁵³ EA s 19(2)(d).

PCP where it has failed to comply with the PSED.⁵⁴ While courts have not yet applied the *Bank Mellat* proportionality test to an AIRAS,⁵⁵ its requirement of ‘rational connection to the objective’ may be a pressure point for litigants, as some AIRAS have been shown to be only marginally more accurate or even less accurate than manual risk assessments.⁵⁶ Claimants may also rely on the importance of the interests affected by a decision—and correspondingly the ‘seriousness of the detriment’⁵⁷ caused by the discrimination—to argue that the decision did not strike a fair balance between individual rights and community interests. Moreover, the level of scrutiny to which courts subject government justifications may be heightened if judges have regard to the academic criticism of discriminatory AIRAS indicating that such systems risk entrenching systemic injustice as well as disadvantaging individuals.⁵⁸ What remains uncertain is how these factors will be weighed against submissions by public authorities emphasising the significant public benefit created by the cost effectiveness and efficiency of AIRAS-assisted decision-making.

II. Evidential Challenges

Algorithmic opacity creates evidential challenges for claimants in establishing that an AIRAS is discriminatory. These challenges are especially acute for claimants alleging direct discrimination, as in the context of a complex or privately owned machine learning AIRAS it will be near-impossible to positively establish that the

⁵⁴ ‘Statutory Code of Practice’ (n 46) 81. See for example *R(Ward) v Hillingdon LBC* [2019] EWCA Civ 692 [69].

⁵⁵ *Bank Mellat v HM Treasury (No 2)* [2013] UKSC 39.

⁵⁶ Julia Dressel and Hany Farid, ‘The accuracy, fairness and limits of predicting recidivism’ (2018) 4 *Science Advances*.

⁵⁷ *R (Elias) v Secretary of State for Defence* [2006] IRLR 934 [151].

⁵⁸ For example see Oswald et al. (n 24) 228; Veale et al. (n 6) 18.

reason for an individual's less favourable treatment is membership of a protected group.

Tomlinson, Sheridan and Harkens have argued that courts should take a 'distinct approach' to the standard of evidence in judicial review of opaque automated systems, by either adjusting the standard depending on the degree of opacity in a system, or shifting the duty of explaining how a system works to the public authority.⁵⁹ However, they concede that principles of evidence in judicial review are difficult to distill, lying in the 'legal subconscious',⁶⁰ and there is no guarantee that courts would choose to adopt such an approach.

However, claimants may be aided by established equality law principles applied to the automated decision-making context. EA 2010 subsections 136(2) and (3) reverse the burden of proof where a claimant successfully establishes facts from which the court could decide that discrimination had occurred. In *ex p Danfoss*,⁶¹ the European Court of Justice (ECJ) held that a lack of transparency may give rise to an inference of direct discrimination where less favourable treatment is established. In that case, the claimants alleged sex-based pay discrimination, relying on a statistical disparity in average pay; the Court found this sufficient to place the burden on the employer, since it was not possible for employees to identify the criteria used to determine pay. Allen and Masters consider this equality law principle to be generally applicable to public functions.⁶² While ECJ jurisprudence has never been binding in UK judicial proceedings unless the public

⁵⁹ Joe Tomlinson, Katy Sheridan, and Adam Harkens, 'Proving Public Law Error in Automated Decision-Making Systems' (PLP Annual Conference, London, October 2019) 13.

⁶⁰ *ibid* 3.

⁶¹ Case C-109/88 [1991] ICR 74.

⁶² Allen and Masters (n 41) 51.

authority was acting to fulfil EU law obligations, and its status was recently further diminished by the end of the Brexit transition period,⁶³ much of the EA 2010 is nevertheless derived from the UK's EU law obligations. Applying the principle that 'it is inconceivable that Parliament intended the same concepts to be interpreted differently in different contexts',⁶⁴ courts adjudicating an AIRAS may draw a similar inference to that in *Danfoss*. It is therefore plausible that a claimant seeking judicial review of an opaque AIRAS could persuade a court to infer direct discrimination so long as they were able to identify a statistical disparity in outcome between members of a protected group and others,⁶⁵ placing the burden of proof on the public authority; where a public authority chooses to maintain intentional opacity or where the system is intrinsically opaque, it will be unable to discharge this burden.

Claimants in indirect discrimination cases would not need to establish a causal link between their membership of a protected group and the disadvantage suffered, and so would be able to avail themselves of the section 136 presumption by merely demonstrating a statistical disparity in outcomes. Though these claimants would not need to rely on the *Danfoss* inference, the level of opacity present within a system would nevertheless be relevant to the court's assessment. In order to rebut the presumption of discrimination, the public authority would need to establish either that there was no causal link between the PCP

⁶³ See for example David Feldman, 'Departing from Retained EU Case Law' (*UK Constitutional Law Association Blog*, 11 January 2021) <<https://ukconstitutionallaw.org/2021/01/11/david-feldman-departing-from-retained-eu-case-law/>> accessed 5 April 2021.

⁶⁴ *Essop* (n 50) [19] (Lady Hale).

⁶⁵ Indeed, there has been increased focus by courts on the use of quantitative data to prove error: Tomlinson, Sheridan and Harkens (n 59) 6, referring to *R (UNISON) v Lord Chancellor* [2017] UKSC 51.

and the disadvantage suffered by the individual, or, if it failed to do so, that the measure is justified. It would be very difficult to demonstrate that a PCP satisfies a proportionality test where a public authority itself is unable to comprehend its workings, since, for example, the public authority will be unable to establish that no less discriminatory measures would have satisfied the desired objective. Moreover, complete opacity will, as discussed below, make it difficult for the authority to satisfy its PSED which, in turn, is likely to harm the authority's ability to justify the measure.

B. The PSED

I. The Potential of the PSED to Eliminate Opacity

The PSED, properly harnessed, has significant potential to prevent discriminatory AIRAS decisions and to counter opacity. Under EA 2010 section 149(1), public authorities have a mandatory duty to have 'due regard' to three equality issues when exercising their functions.⁶⁶ The duty applies to all 'functions' of public authorities, so that both policy and individual decisions are subject to challenge;⁶⁷ in respect of AIRAS, challenges may be made to the procedure followed in making a high-level decision to introduce a tool into decision-making, but also to the procedural element of individual decisions made based on risk scores.

⁶⁶ The three issues are the need to: (a) eliminate discrimination and other conduct prohibited by the Act; (b) advance equality of opportunity between those who share a relevant protected characteristic and those who do not; and (c) foster good relations between those who share a relevant protected characteristic and those who do not.

⁶⁷ *Pieretti v Enfield LBC* [2010] EWCA Civ 1104 [26].

Though subsections 149(3)-(5) elaborate the terms of these three duties in detail and they are conceptually distinct from one another, the jurisprudence has largely conflated them to focus on the more general question of whether there has been adequate consideration of the negative equality impacts of decisions.⁶⁸ While this is generally a question for the court,⁶⁹ the *weight* to be given to the relevant section 149(1) ‘needs’ is a matter for the decision-maker, subject only to rationality review.⁷⁰ A successful PSED challenge will render a decision void for illegality and will have the added effect of lending support to a concurrent indirect discrimination claim.

Although the PSED is a procedural rather than substantive obligation and decision-makers are not prevented from simply making the same decision again after it has been quashed, the duty may produce substantive outcomes indirectly due to its potential to combat algorithmic opacity. Hickman notes that because compliance is a fact-specific question, the courts give ‘very close scrutiny’ to questions of fact in PSED cases.⁷¹ In judicial review proceedings, reliance is ordinarily placed on the public authority’s duty of candour, so that proceedings are conducted on the basis that a decision-maker’s witness statement about the basis for a decision is accurate. In PSED cases, however, the courts have been willing to closely review primary evidence—including internal documentation—and reject the views of public authorities,⁷² resulting in an ‘onerous’ burden on

⁶⁸ Tom Hickman, ‘Too hot, too cold or just right? The development of the public sector equality duties in administrative law’ [2013] Public Law 325, 326.

⁶⁹ R (*JM*) v *Isle of Wight Council* [2011] EWHC 2911 (Admin) [102].

⁷⁰ R (*Harris*) v *Haringey LBC* [2010] EWCA Civ 703 [40].

⁷¹ Hickman (n 68) 339 notes the example of R (*JL*) v *Islington LBC* [2009] EWHC 458 (Admin).

⁷² Hickman (n 68).

authorities ‘to demonstrate that all equality issues have been considered and to close off all arguments [about non-compliance]’.⁷³ This high evidential burden has effects both at the decision-making stage and at the litigation stage.

At the point of decision-making, the duty incentivises public authorities to conduct investigations about discriminatory effects of AIRAS and record their findings. Indeed, Aikens LJ noted in *R (Brown) v Secretary of State for Work and Pensions* that it was good practice to keep ‘adequate record’ showing that the PSED had been ‘actually considered’.⁷⁴ Moreover, the PSED incentivises public authorities to avoid systems with high levels of intrinsic or illiterate opacity, as they will be unable to demonstrate compliance where they are themselves unable to understand an AIRAS’ equality impacts. It has also been suggested by practitioners that courts may be amenable to an argument that public bodies cannot comply with their PSED unless they disclose details of automated systems for an independent assessment.⁷⁵

Moreover, the evidence provided by the public authority at trial in the course of PSED claims aids in creating transparency about a given system, regardless of the ultimate success of the claim. This will not only incentivise decision-makers to eliminate discrimination *ex ante*, but may act as a valuable tool to assist litigants in gathering evidence of discrimination to support subsequent or concurrent judicial review challenges on other grounds such as illegality (under EA 2010 section 29 or the Human Rights Act 1998), or under common law grounds including irrationality and the doctrine of relevant considerations. This may be the case even where permission is refused, as

⁷³ *ibid* 340-341.

⁷⁴ [2008] EWHC 3158 (Admin) [96].

⁷⁵ See comments by Megan Goulding in Irving (n 15).

evidenced by *R (Fawcett Society) v Chancellor of the Exchequer*,⁷⁶ in which permission to apply for judicial review of the national budget was refused only after a ‘substantial hearing and... detailed judgment’.⁷⁷

II. The Decision in Bridges

The potential of the PSED in combating algorithmic opacity may be illustrated by the decision of the Court of Appeal in *R (Bridges) v Chief Constable of South Wales Police*.⁷⁸ The case concerned a challenge to the South Wales Police’s (SWP’s) use of automated facial recognition (AFR) technology on the grounds that, *inter alia*, the SWP had failed to comply with the PSED as it had not considered the possibility that AFR was indirectly discriminatory in producing a higher rate of false positives for women and Black and minority ethnic people.

The Divisional Court found that there was no violation since there was ‘no firm evidence’ of discriminatory results⁷⁹ and ‘no specific reason’⁸⁰ for the SWP to believe there might be discriminatory effects when it commenced use of AFR. Strikingly, the Court reached this conclusion despite the statement of the claimant’s expert witness, who noted that bias was a common feature of AFR systems⁸¹ but explained he could reach no definite conclusion on SWP’s system because he did not have access to the training data—the claimant’s request for access was refused on the basis that the data was a trade secret⁸²—and that the

⁷⁶ [2010] EWHC 3522 (Admin).

⁷⁷ Hickman (n 68) 333.

⁷⁸ [2020] EWCA Civ 1058.

⁷⁹ *R (Bridges) v Chief Constable of South Wales Police* [2019] EWHC 2341 (Admin) [153].

⁸⁰ *ibid* [157].

⁸¹ *ibid* [155].

⁸² See Irving (n 15).

defendant, who also had no access, could not make the evaluation either.

However, the Court of Appeal took a more expansive view of the PSED as mandating a public authority to mitigate algorithmic opacity rather than being itself impeded by it. Allowing the appeal, the Court of Appeal found that the Divisional Court's approach had 'put the cart before the horse', since 'the whole purpose of the positive duty... is to ensure that a public authority does not inadvertently overlook information which it should take into account'.⁸³ Though the Court conceded that commercial confidentiality 'may be understandable' it could not 'enable a public authority to discharge its own, non-delegable, duty under section 149'.⁸⁴ In the Court's assessment, the SWP should have 'sought to satisfy themselves, either directly or by way of independent verification, that the software program in this case does not have an unacceptable bias on grounds of race or sex'.⁸⁵ The Court also explicitly recognised the importance of combating opacity for *both* the procedural PSED and the substantive prohibitions against direct and indirect discrimination under the EA 2010.⁸⁶

The Court of Appeal's decision offers some guidance to public authorities on how they may comply with the PSED even where systems are opaque, though it remains to be seen what arrangements with third parties courts will consider sufficient to constitute 'independent verification'. What appears certain, however, is that reliance on a developer's bare assurance that a system is not biased will not meet the threshold. It is hoped that

⁸³ R (*Bridges*) v *Chief Constable of South Wales Police* [2020] EWCA Civ 1058 [182].

⁸⁴ *ibid* [199].

⁸⁵ *ibid* [199].

⁸⁶ *ibid* [200].

the Court of Appeal's approach will be applied appropriately by first instance courts considering AIRAS, as well as other automated decision-making and decision-assistance systems, and that future appellate decisions will continue to utilise the PSED to facilitate transparency and accountability.

Conclusion

Despite the relative novelty of AIRAS and the challenges of opacity, it is evident upon closer examination that the existing framework of the EA 2010 may prove to be an unexpectedly effective avenue for challenging discriminatory systems. This may extend to influencing government decision-making before litigation takes place, as is evident from the Home Office's discontinuation of the use of its rules-based visa 'streaming' tool after the The Joint Council for the Welfare of Immigrants and the charity Foxglove obtained permission to apply for judicial review of its use.⁸⁷ The claimants alternatively submitted that the use of nationality⁸⁸ as a criterion in streaming individuals into risk categories⁸⁹ was directly discriminatory, that other criteria used

⁸⁷ David Bolt, 'An inspection of entry clearance processing operations in Croydon and Istanbul – November 2016-March 2017' (2017) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/631520/An-inspection-of-entry-clearance-processing-operations-in-Croydon-and-Istanbul1.pdf> accessed 5 April 2021.

⁸⁸ See EA 2010 s 9(1)(b).

⁸⁹ See Rafe Jennings, 'Government Scraps Immigration 'Streaming Tool' before Judicial Review' (*UK Human Rights Blog*, 6 August 2020) <<https://ukhumanrightsblog.com/2020/08/06/government-scraps->

were directly or indirectly discriminatory, and that the Home Office had violated its PSED.⁹⁰ The Home Office subsequently agreed to pause use of the tool while it was redesigned,⁹¹ an outcome which further evidences the effectiveness of the EA 2010 in rising to the equality challenges posed by novel forms of technology-facilitated bias.

However, this article does not seek to make the argument that piecemeal development of principles through *ex-post* review under the EA 2010, or Government reversals caused by the threat of litigation, is sufficient to counter the challenges of algorithmic bias in AIRAS. This is firstly because the EA 2010 can offer protection only against forms of bias which constitute discrimination under its narrow definition; it does not, for example, prohibit bias on the basis of socio-economic class, and the procedural socio-economic duty in section 1 has not been commenced. Other grounds of review, such as those noted above at Part 3.B.I, also do not cover every instance of algorithmic bias. More importantly, bias in AIRAS is not merely a challenge for administrative law, but also raises important policy questions which should be addressed by democratic decision-making. Decisions need to be made as to the types of decision, if any, that AIRAS are suitable for, the degree of acceptable bias (and the groups against which bias is acceptable), as well as the need for

immigration-streaming-tool-before-judicial-review/> accessed 5 April 2021.

⁹⁰ As noted in the letter before claim sent to the Home Office.

⁹¹ Home Office statement quoted in Henry McDonald, 'Home Office to scrap 'racist algorithm' for UK visa claimants' *The Guardian* (London, 4 August 2020) <www.theguardian.com/uk-news/2020/aug/04/home-office-to-scrap-racist-algorithm-for-uk-visa-applicants> accessed 5 April 2021.

independent assessment or monitoring systems.⁹² Politically driven reform could also offer a more comprehensive framework to promote *ex-ante* accountability, such as through the creation of a regulatory body or, as proposed by Cartwright in this Journal, a regulatory ‘toolkit’.⁹³

However, unless the Government demonstrates appetite for reform in this area,⁹⁴ judicial review will likely remain the main mechanism for combating discrimination in AIRAS for some time. Accordingly, it is worth continuing to follow developments on how courts will choose to apply established equality law principles to emerging technologies.

⁹² For an example of political consideration of bias in AIRAS, see the European Commission’s recent proposal for a Regulation introducing a legal framework on artificial intelligence, available at: <<https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-approach-artificial-intelligence>> accessed 21 April 2021. The draft Regulation subjects certain types of ‘high-risk’ artificial intelligence systems to special requirements in respect of training data, including requiring ‘examination in view of possible biases’: article 10(2)(f). ‘High-risk’ systems include those ‘intended to be used by law enforcement authorities for making individual risk assessments of natural persons in order to assess the risk of a natural person for offending or reoffending’: Annex III.

⁹³ Benjamin Cartwright, ‘Regulating the Robot: A Toolkit for Public Sector Automated Decision-Making’ (2021) 10 OUULJ 21.

⁹⁴ The Government could choose to take forward reform internally or to commission work or to refer the issue for consideration by an independent body. For example, the Law Commission has recently invited comment on the potential inclusion of a project on automated decision-making in its 14th Programme of Law Reform—see <<https://www.lawcom.gov.uk/14th-programme-kite-flying-document/>> accessed 5 April 2021.

Male Mothers: The Law's Struggle to Capture the Modern Family

Caragh Deery*

Abstract—The Court of Appeal in *R (McConnell and YY) v Registrar General* upheld a High Court ruling that motherhood is defined by the role a person plays in the biological process of conception, pregnancy and birth regardless of whether that person is legally considered to be a man or a woman. By concluding that section 12 of the Gender Recognition Act 2004 means that a person's status as the legal mother or father of a child is unaffected by his or her legally acquired gender,¹ the Court detached motherhood from gender. This had the effect that Mr McConnell, a transgender man, had to be registered as the 'mother' of his child (YY) despite the fact that this did not match the social reality of his role as YY's 'father'. This paper examines the decisions of the High Court and the Court of Appeal and addresses the criticisms raised against them. It seeks to show that

* St Hilda's College, Oxford. I am very grateful to Rachel Taylor, Julie Summers, and the OUULJ editorial team for their comments on earlier drafts. Any errors remain my own.

¹ The term 'legally acquired gender' is used here to describe the legal process by which a transgender person's gender is confirmed through obtaining a Gender Recognition Certificate under the provisions of the Gender Recognition Act 2004.

McConnell was doctrinally sound and is consistent with other areas of family law, but that the case is demonstrative of wider public policy issues. The issues raised in *McConnell* reflect a pressing need for a thorough review of the law in this context.

Introduction

This paper discusses the recent High Court² and Court of Appeal (CoA)³ decisions in *R (McConnell and YY) v Registrar General*. The proceedings were the culmination of a transgender man's plea to be legally recognised as his child's 'father', or otherwise 'parent' or 'gestational parent'. The CoA, upholding the High Court decision, refused and held that Mr McConnell could only be registered as his child's 'mother'. It reached this conclusion by undertaking a proportionality assessment and determining that Parliament had been within its margin of discretion to prioritise the right of a child born to a transgender parent to know the biological reality of its birth, rather than the right of the transgender parent to be registered in their confirmed gender. The United Kingdom Supreme Court (UKSC) refused Mr McConnell's application to appeal on the basis that the application did not raise an arguable point of law that ought to be considered at that time.⁴ This marked the end of Mr McConnell's legal journey—at least in the UK courts—but widespread doubts remain as to whether justice has been achieved.

The author argues that the decision in *McConnell* was doctrinally sound, but that it reflects broader public policy failures of family law to capture adequately modern and diverse family forms. The judgment was consistent with other areas of family law and struck a proportionate balance between the competing rights at stake. Moreover, the CoA afforded sufficient respect to the competence of Parliament by recognising that its ability to

² *R (IT) v Registrar General for England and Wales (AIRE Centre intervening)* [2019] EWHC 2384 (Fam), [2020] Fam 45.

³ *R (McConnell and YY) v Registrar General* [2020] EWCA Civ 559, [2020] 2 All ER 813.

⁴ *R (McConnell and YY) v Registrar General* UKSC 2020/0092.

initiate reform was restricted by its own expertise. Transformative change of the law in such a socially and politically sensitive context is a job for Parliament and not something that could have been achieved through a tenuous interpretation of the law. If the Court had taken the route argued for by Mr McConnell and allowed him to be registered as YY's 'father' or 'parent', its decision would only accommodate the individual interests of parties who find themselves in analogous positions to the relatively narrow set of circumstances before the Court. This would have widespread and potentially grave implications for the rights of others and would cause uncertainty for other statutes reliant on the word 'mother' to allocate legal rights or duties. For these reasons, the Court was correct to prioritise the rights of YY and children generally to ascertain definitively their biological origins.

Nonetheless, the fact that the law requires Mr McConnell to be registered as YY's 'mother' in the first place is problematic. The decision fails to reflect social reality and demonstrates a clear need for the Gender Recognition Act 2004 (GRA 2004) to be reviewed by Parliament. *McConnell* highlights the need to consider whether reform to better reflect social reality is possible and raises the question of whether the time has come to de-gender parenthood.

Devising a coherent birth registration scheme that consistently registers the person who gave birth to a child without relying on the term 'mother' is perfectly feasible and is a step which has already been taken in other jurisdictions.⁵ The CoA was nonetheless correct to adopt a cautious approach and to refuse to issue a declaration of incompatibility under section 4 of the

⁵ See discussion on the Ontarian All Families Are Equal Act in Section 5.

Human Rights Act 1998 (HRA 1998). The risks are patent. The political staying power of judicial declarations of incompatibility runs the risk of clouding future political discourse at the expense of other interests. De-gendering parenthood would require a careful and thorough review of the law. The dangers identified by the CoA—namely the implications for the rights of others and the impacts upon existing statutory schemes—must remain at the forefront of Parliamentary debate.

In making these arguments, this paper adopts the following structure. Section 1 outlines the societal and legal developments that led to the culmination of the issues in *McConnell*. Section 2 provides the facts of the case. Section 3 analyses the High Court and CoA judgments. Section 4 examines the criticisms levied against *McConnell* and considers whether the CoA should have made a declaration of incompatibility under section 4 HRA 1998. Finally, Section 5 contextualises *McConnell* within family law more broadly and considers the path forward.

1. *The Law's 'Fear' of Pregnant Men*

The litigation in *McConnell* was unimaginable fifty years ago. Previously, the position under English law was that a person's sex was determined at birth and was unchangeable.⁶ In a time before artificial reproductive technology (ART), and when transgender individuals were unrecognised by the law, the determination of motherhood was so obvious as not to warrant judicial or

⁶ *Corbett v Corbett (otherwise Ashley)* [1971] P 83; Mary Welstead, 'Every Child Should Have a Mother' [2020] Fam Law 1099, 1102.

legislative attention.⁷ Developments towards greater rights for transgender persons led to legal and regulatory difficulties. There was a public policy concern to prevent procreative practices considered undesirable and problematic by the law.⁸ Margaria described the ‘fear’ that ‘pregnant men’ would destabilise Europe’s family law systems and confuse children about their biological origins, depriving them of important family relationships.⁹ Until recently, most jurisdictions required persons wishing to be legally recognised by another gender to undergo gender confirmation surgery and/or sterilisation to prevent such scenarios from arising.

The European Court of Human Rights’ (ECtHR’s) decision in *Goodwin* was the turning point.¹⁰ It held that presumptions that sex was unchangeable violated Article 8 of the European Convention on Human Rights (ECHR).¹¹ The UK Parliament subsequently enacted the GRA 2004 to ensure compliance. Going a step further than the ECtHR required of it, it gave legal recognition to transgender persons whether or not they had undergone gender reassignment surgery. In any event, such requirements were challenged directly in 2017 when the ECtHR found a French law demanding sterilisation of transgender men incompatible with Article 8.¹² These

⁷ Claire Fenton-Glynn, ‘Deconstructing Parenthood: What Makes a “Mother”?’ (2020) 79 *Cambridge Law Journal* 34, 34.

⁸ Alice Margaria, ‘Trans Men Giving Birth and Reflections on Fatherhood: What to Expect?’ (2020) 34 *International Journal of Law, Policy and the Family* 225, 229.

⁹ *ibid* 232; Peter Dunne, ‘Transgender Sterilisation Requirements in Europe’ (2017) 25 *Medical Law Review* 554, 564.

¹⁰ *Christine Goodwin v UK* (2002) 35 EHRR 447.

¹¹ Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights, as amended) (ECHR) art 8.

¹² *AP, Garçon and Nicot v France* (2017) ECHR 121.

developments mean it is no longer possible for Convention States to avoid addressing the legal status of ‘pregnant men’. As McFarlane P acknowledged in *McConnell*, ‘...[i]t is now medically and legally possible for an individual, whose gender is recognised in law as male, to become pregnant and give birth to their child.’¹³ Issues relating to social and legal definitions of ‘mothers’ and ‘fathers’ can no longer be ignored, and it was the courts which were faced with the unenviable task in *McConnell*.

2. *The Facts of McConnell*

Mr McConnell was registered female at birth but transitioned to live in the male gender around a decade before the proceedings before the CoA. He had taken steps to medically confirm his gender by undergoing testosterone therapy and a double mastectomy, and his passport and NHS records were amended to reflect his male gender. In 2016, Mr McConnell wished to begin a family and suspended testosterone treatment to commence fertility treatment.

In 2017, Mr McConnell issued an application under the GRA 2004 to obtain a Gender Recognition Certificate (GRC) confirming that he was male. The effect of a GRC is that the applicant ‘becomes for all purposes the acquired gender’,¹⁴ but, as will be shown, this is subject to exceptions.¹⁵ An applicant must provide medical reports confirming a diagnosis of gender dysphoria and a declaration that the applicant intends to continue

¹³ *McConnell* EWHC (n 2) [279].

¹⁴ Gender Recognition Act 2004, s 9(1).

¹⁵ *ibid* s 9(3).

to live in the acquired gender until death. As aforementioned, an individual does not need to undergo sterilisation or gender reassignment surgery to be legally recognised as their confirmed gender in the UK. This meant that Mr McConnell could obtain a GRC and be legally recognised as a male, notwithstanding that he retained the reproductive organs of a biological woman.

Ten days after the issuance of his GRC, Mr McConnell underwent intrauterine insemination (IUI) fertility treatment. Donor sperm was placed inside his uterus in a successful attempt to fertilise one or more of his eggs. Mr McConnell became pregnant and later gave birth to a son (YY) in 2018. When registering YY's birth, Mr McConnell was informed by the Registry Office that he had to be registered as the child's 'mother' even though the registration could be in his current (male) name.

Mr McConnell brought proceedings for judicial review of the Registry's decision, seeking a declaration that, as a matter of law, he was to be regarded as YY's 'father', 'parent' or 'gestational parent' and was entitled to be registered accordingly. Alternatively, he alleged that if the law required him to be registered as YY's 'mother', he would be entitled to a declaration of incompatibility under s 4 HRA 1998. He argued that registering him as YY's 'mother' would infringe his and/or YY's rights under Article 8 (the right to respect for private and family life) and Article 14 (the right to enjoy the rights and freedoms of the Convention free from discrimination) of the ECHR.

3. The Decisions of the High Court and the Court of Appeal

The proceedings first appeared before the President of the Family Division of the High Court. McFarlane P determined that, in law, Mr McConnell was YY's 'mother' for the purposes of the registration of YY's birth.¹⁶ Dissatisfied with this outcome, Mr McConnell brought his case to the CoA. The panel, headed by Lord Burnett, dismissed the appeals and upheld the decision of the High Court. By concluding that s 12 GRA 2004 means that a person's status as the legal mother or father of a child is unaffected by his or her legally acquired gender, the courts detached motherhood from gender. Both courts acknowledged that this amounted to an interference with the appellants' Article 8 rights but ultimately found this to be justified.

This paper divides its analysis of the decisions into three stages. Firstly, it considers the definition afforded to motherhood and the subsequent implications for the correct interpretation of ss 9 and 12 GRA 2004. Secondly, it examines the courts' analysis as to whether that definition infringed the appellants' Article 8 rights. Thirdly, it analyses the CoA's proportionality assessment in determining that those provisions are not incompatible with the ECHR.

A. The Interpretation of Motherhood

A preliminary consideration that had to be addressed by both the High Court and the CoA was the legal definition of motherhood. The courts were rightly cautious to ensure consistency with other areas of family law.

¹⁶ *McConnell* EWHC (n 2) [282].

I. The Courts' Interpretation of Motherhood and its Application to the GRA 2004

McFarlane P explained that it is the role of gestating and giving birth that is 'at the essence of what a 'mother' undertakes with respect to a child, under the common law position.'¹⁷ This focuses on the 'biological process' resulting in the birth of a child 'rather than the person's particular sex or gender'.¹⁸ In contrast, and perhaps surprisingly, the CoA largely put aside the meaning of 'mother' at common law and in registration legislation. A working definition of what it means to be a 'mother' or 'father' was a necessary first step to determine the correct interpretation of ss 9 and 12 GRA 2004. By sidestepping this discussion, the High Court's common law conception of motherhood must be taken to have been implicitly accepted by the CoA.¹⁹

Mr McConnell disagreed with the common law definition of motherhood. He argued that developments in statutory law, such as the Human Fertilisation and Embryology Act 2008 (HFE Act) and the GRA 2004 meant the common law position was no longer absolute. Section 12 of the GRA 2004 indicates otherwise. It explicitly states that 'the fact that a person's gender has become the acquired gender under this Act does not affect the status of the person as the father or mother of a child'. Mr McConnell maintained that this exception to s 9(1) GRA 2004—which sets out that the holder of a GRC 'becomes for all purposes the acquired gender'—should only apply retrospectively in relation to children conceived before the parent acquired his or her GRC. He argued that this interpretation was necessary to

¹⁷ *ibid* [139].

¹⁸ *ibid* [139].

¹⁹ *McConnell* EWCA (n 3) [35].

avoid leaving individuals ‘in limbo between two genders.’²⁰ If the provision applied prospectively, it would render a GRC effective in determining gender for all purposes other than parenthood and this would defeat the purposes of the GRA 2004.

McFarlane P rejected this interpretation by once more emphasising the common law position, maintaining that Mr McConnell’s argument would only have force if the attribution of the status of ‘mother’ or ‘father’ was gender specific.²¹ The CoA put forward several further convincing reasons that indicated that s 9(1) was both prospective and retrospective in its effect.²² First, it was not, on the face of it, limited to events occurring before the issuance of a GRC. Second, s 9(2), which states that a GRC does not affect events that had already happened, would be otiose if s 12 only had retrospective effect. Third, the wording of s 12 was similar to that used in other sections of the Act which marked out exceptions to s 9(1). Finally, other provisions in the Act which were only intended to have retrospective effect were drafted with express language to that effect. The CoA concluded that Parliament intended the provision to be both prospective and retrospective, with the effect that it captured Mr McConnell. For the purposes of the registration of YY’s birth, Mr McConnell was YY’s ‘mother’.²³

II. Criticisms of the Courts’ Interpretation

The CoA’s approach to motherhood in the context of the GRA 2004 has been criticised on the basis that it disadvantages transgender parents. Spokespersons for Stonewall, an LGBTQ+ rights charity, describe the judgment as ‘another example of how

²⁰ *McConnell* EWHC (n 2) [87].

²¹ *ibid* [145].

²² *McConnell* EWCA (n 3) [29]-[33].

²³ *ibid* [39], [89].

current legislation contradicts the fragile equality trans people currently have' and as one that fails to recognise transgender people for who they are.²⁴ Margaria also criticises the decision's failure to reflect social reality by producing a contradictory outcome whereby children can have a 'father in life, but a mother in law'.²⁵

However, such criticisms overlook that the CoA was neither attempting, nor facilitating, the exclusion of transgender parents from the same rights enjoyed by cisgender parents. Rather, the Court's interpretation of 'motherhood' was partly driven by a concern to ensure consistency with other areas of family law. The emphasis placed on the role of gestating and giving birth to a child in the determination of motherhood is uniform across English family law,²⁶ with the result that the disjunction between the social and legal realities of parenthood exists in other contexts. For example, the CoA referred to s 33(1) of the HFE Act 2008 which, in the context of IVF, stipulates that '[t]he woman who is carrying or who has carried a child... and no other woman, is to be treated as the mother of the child.'²⁷ In the context of surrogacy, Baroness Hale explains that the surrogate (i.e. gestating)²⁸ mother will always be the 'legal parent unless and until a court order is made in favour of the commissioning parents.'²⁹ For adoption, the policy choice made by Parliament is

²⁴ Laura Russel, 'Statement on the Ruling Against Freddy McConnell' (*Stonewall*, 29 April 2020).

<<https://www.stonewall.org.uk/about-us/news/statement-ruling-against-freddy-mcconnell>> accessed 10 December 2020.

²⁵ Margaria (n 8) 227.

²⁶ *McConnell* EWCA (n 3) [66]-[71].

²⁷ *ibid* [67].

²⁸ The Surrogacy Arrangements Act 1985, s 1(2) defines the surrogate mother as the woman who carries the child.

²⁹ *XX v Whittington Hospital NHS Trust* [2020] UKSC 14, [2020] 4 All ER 93 [9].

once again that the person who gives birth to the child is the only mother.³⁰ This is reflected on the birth certificate even though, like in *McConnell*, this might not reflect social reality. For the same reason that Parliament believes an adopted child should be able to look at its birth certificate and learn that it was adopted by its social parents, the child of a transgender parent should be able to look at its birth certificate to learn that its social ‘father’ is its biological ‘mother’. From this perspective, the CoA’s reasoning appropriately safeguards the doctrinal coherence of the law.

B. Were the Appellants’ Rights Infringed?

Both the High Court and the CoA found that there had been an interference with the appellants’ Article 8 rights. McFarlane P acknowledged that, particularly from Mr McConnell’s perspective, the degree of interference was substantial.³¹ Whilst it is true that legal parenthood fails to reflect social parenthood in other contexts, the impact on Article 8 rights is undoubtedly starker where that disjunction also fails to reflect an individual’s confirmed gender. Gender is, for many, a central tenet of identity and the transgender community has fought hard for legal recognition. Although the legal conception of ‘motherhood’ might now be considered to be detached from gender, the terms ‘mother’ and ‘father’ are widely understood in gendered terms and thus go to the very heart of gender dysphoria.³² The current legal position might have the result that individuals like Mr McConnell will be forced to choose either to have a family, and be ‘outed’ as transgender (albeit in limited circumstances), or to abandon the prospect of parenthood in order to retain their acquired gender

³⁰ *McConnell* EWCA (n 3) [71].

³¹ *McConnell* EWHC (n 2) [272].

³² Fenton-Glynn (n 7) 36.

for all purposes.³³ To require him to be registered as YY's 'mother' was therefore rightly seen by Mr McConnell as a frontal assault on the integrity of his legally acquired male gender that would adversely impact his human dignity.³⁴

Whilst the conclusion that there had been an infringement might seem uncontroversial, it has been challenged. Welstead wonders whether the Court went too far in determining that Mr McConnell's rights had been infringed *at all*.³⁵ He had been 'less than honest' in his declaration to the Gender Recognition Panel that he intended to live as a male for the remainder of his life when he planned, within a matter of days, to undergo IUI.³⁶ To this end, McFarlane P expressed two concerns of his own. First, there was the possibility that Mr McConnell's conduct might have constituted fraud.³⁷ Second, a further concern was raised that the clinic treating Mr McConnell might have breached its license by providing him with fertility treatment. The HFEA legislation currently only permits the HFEA to license a clinic to provide services to assist 'women to carry children'.³⁸ Neither point was at issue before the courts and McFarlane P did not decide on either point. Nevertheless, Welstead maintains that if Mr McConnell had revealed his reproductive intention, he would not have been granted a GRC and his 'problematic wish' to register YY's birth in his acquired gender would not have arisen in the first place.³⁹

³³ *ibid.*

³⁴ *McConnell* EWHC (n 2) [251].

³⁵ Welstead (n 6) 1106.

³⁶ *ibid.*

³⁷ *McConnell* EWHC (n 2) [45].

³⁸ Human Fertilisation and Embryology Act 1990, s 2.

³⁹ Welstead (n 6) 1106.

This position is dismissive of Mr McConnell's fundamental rights and the ongoing struggle for transgender equality more broadly. Welstead fails to recognise that the real problem is that the current legal rules require some transgender men to resort to such decisions to feel as though their rights are protected, and their confirmed gender fully recognised in law. In any event, the idea that Mr McConnell was acting 'dishonestly' is based on the outdated idea that living in the male gender and becoming pregnant are mutually exclusive concepts. The medical and legal reality is such that this can no longer hold. Any suggestion that Mr McConnell's Article 8 rights were unaffected is thus untenable.

4. The Proportionality Analysis

A. Fair Balance

As regards a 'fair balance', the CoA's analysis was coherent and struck an appropriate balance between the relevant interests at stake in *McConnell*. Article 8(2) ECHR provides exceptions under which it is permissible for a public authority to interfere with the right to private and family life in Article 8(1). One example is where interference is necessary for the protection of the rights and freedoms of others. The Court correctly placed significant emphasis on the rights of YY and other children who would be affected by the decision generally, referring to Article 3(1) of the United Nations Convention on the Rights of the Child (CRC). This provides that, in all actions concerning children 'the best interests of the child shall be a primary consideration.'⁴⁰ The

⁴⁰ *McConnell* EWCA (n 3) [83].

Court noted that although the CRC has not been incorporated into domestic law by Parliament, both the ECtHR and the domestic courts have regard to it when interpreting Article 8 ECHR.⁴¹

The importance of this right for children to know their biological origins should not be underestimated. As Munby P famously asserted in *Re HFEA*, legal parenthood is:

[A] question of most fundamental gravity and importance. What, after all, to any child, to any parent, never mind to future generations and, indeed, to society at large, can be more important, emotionally, psychologically, socially and legally, than the answer to the question: who is my parent? Is this my child?⁴²

The author argues that the Court was correct to prioritise this right over any Article 8 right which YY (or Mr McConnell) would have enjoyed as a result of Mr McConnell being registered as the legal ‘father’ or ‘parent’ of YY. This is particularly true because although registering Mr McConnell as YY’s ‘mother’ would fail to reflect the social reality of the parent-child relationship,⁴³ that relationship would be largely unaffected in practical terms. The registration of Mr McConnell as YY’s ‘mother’ would only be recorded on the long-form version of the birth certificate which would only need to be produced in a limited set of circumstances,⁴⁴ whereas the right to biological certainty is fundamental to individual identity. This was a crucial

⁴¹ *ibid* [85].

⁴² *Re A and other cases (Human Fertilisation and Embryology Act 2008) (assisted reproduction: parent)* [2015] EWHC 2602 (Fam), [2016] 1 All ER 273 [3].

⁴³ *McConnell* EWCA (n 3) [55].

⁴⁴ *ibid*.

justification for the Court's decision to interfere with Mr McConnell's Convention rights.

Adducing further support for the idea that the judgment was consistent, the right to know one's origins is something which has been emphasised and held capable of compromising competing rights elsewhere in English family law. For example, when ordering a paternity test to be carried out in *Re T (A Child) (DNA Tests: Paternity)*, the Court reiterated the Article 8 right of a child to know one's true parentage.⁴⁵ It held that any interference of the rights of the other parties (including the parents) was proportionate to the legitimate aim of giving the child full knowledge to its parentage.

The rights at stake in the case of parents seeking a paternity test undoubtedly differ in magnitude from the rights of transgender parents seeking to be fully recognised as their confirmed gender. We can, however, observe a near-identical approach to the proportionality analysis of prioritising the right of a child to know its origins in a German case with virtually analogous facts to *McConnell*.⁴⁶ In its judgment, the German Federal Court emphasised the importance of legal parenthood being grounded in biological parenthood to fulfil the legitimate aim to legally assign children to their parents in such a way that is not contrary to biological reality.⁴⁷ It reasoned that if birth registration did not clarify the exact biological reproductive function in which the establishment of the parent-child relationship is grounded, then the child would be deprived of vital

⁴⁵ *Re T (A Child) (DNA Tests: Paternity)* [2001] EWHC Fam 10, [2001] 2 FLR 1190 [57].

⁴⁶ Decision of the Federal High Court of 6 September 2017, BGH XXII ZB 660/14; discussed in *McConnell* EWCA (n 3) [74]-[78].

⁴⁷ *ibid* [9].

information on one's descent.⁴⁸ This demonstrates that the proportionality assessment carried out by the CoA in *McConnell* was not an arbitrary one, as another respected Convention State Court reached the same conclusion when interpreting Article 8 in a factually similar case.

Mr McConnell's individual rights, though important, were therefore outweighed on the basis that both legal certainty and the competing rights of YY required it.

B. 'Less Intrusive Means'

On the facts, the need to respect competing rights and ensure certainty could not readily be achieved in a way that could also protect Mr McConnell's rights. For the appellants to succeed, the CoA would have had to interpret s 12 GRA 2004 contrary to its normal meaning. The Court was conscious of the broader consequences if it were to substitute a word such as 'parent' for the word 'mother' as the appellants had suggested.⁴⁹ A decision which set a precedent that the gestational parent need not be registered as the child's 'mother' would have an impact on the legal status of many child-parent relationships. This is because, as Harrill—a practising family law barrister—has observed, many pieces of interlinked legislation would be affected if the word 'mother' was no longer used to describe the person who gave birth to the child.⁵⁰

The effects of such a substitution would be unjustifiably far-reaching. Take as just one example the fact that the word

⁴⁸ *ibid* [29].

⁴⁹ *McConnell* EWCA (n 3) [65].

⁵⁰ Tom Harrill, 'Modern Families and Assisted Reproduction – Part 1' [2020] *Fam Law* 1198, 1199.

‘mother’ is used 45 times in the Children Act 1989 alone.⁵¹ Section 2(2) of that Act provides potentially the most potent right in the context of parenthood—that a mother has automatic parental responsibility for a child from the moment of birth without any need for registration. This ensures that there is somebody who has legal parental responsibility for every child, and this was considered to be of great importance by the Court.⁵² The ‘less intrusive’ means advocated for by the appellant were therefore inadequate in that they would compromise this right.

Whilst the CoA acknowledged that there was scope for disagreement as to whether the need for every child to have a mother is really in the best interests of children generally, it reasoned that it was not for the Court to reach this conclusion.⁵³ The ‘margin of judgment’ is illuminating here.⁵⁴ The Court was conscious to afford sufficient respect to the judgment of the legislature in determining the issue of justification in relation to the current registration scheme. As aforementioned, the CoA emphasised that Article 3(1) CRC requires the best interests of children *generally* to be taken into account as a primary consideration when striking a balance in legislation.⁵⁵ The Court believed that this is what Parliament intended when enacting a carefully crafted set of provisions in the GRA 2004 which balanced the rights of transgender people and others, including their children.⁵⁶

The Court rightly maintained that its focus was too narrow to make the momentous change argued for by Mr

⁵¹ *McConnell* EWCA (n 3) [64].

⁵² *ibid* [64].

⁵³ *ibid* [86].

⁵⁴ *ibid* [81].

⁵⁵ *ibid* [83].

⁵⁶ *ibid* [86].

McConnell.⁵⁷ The CoA relied on relatively limited evidence, relating to particular circumstances, which was presented by opposing parties. This is woefully inadequate compared to the position of Parliament, which has access to law reform experts, conducts public consultations, and has democratic legitimacy. A Court has no way of determining whether all transgender men would object to being given the status of ‘mother’ after giving birth to a child, nor could it accurately gauge how other members of society would feel if they were no longer referred to as a ‘mother’ or ‘father’ on their child’s birth certificate.⁵⁸ The Court believed that the more acceptable route to achieving transformative reform would be for opponents to take steps to persuade Parliament to take a different view in time.⁵⁹ It was on these bases that the CoA concluded that interference with the appellants’ Convention rights was justified.

5. A Missed Opportunity?

The CoA’s decision has been fiercely criticised on the basis that it ‘missed a vital opportunity’⁶⁰ to challenge the ‘conventional’ understanding of fatherhood⁶¹ which forces transgender people to choose between parenthood and full recognition of their gender. It has rightly been suggested that it would be perfectly feasible to devise a coherent birth registration scheme which consistently registers the person who gave birth to a child without

⁵⁷ *ibid* [81].

⁵⁸ *ibid*.

⁵⁹ *ibid* [86].

⁶⁰ Russel (n 24).

⁶¹ Margaria (n 8) 235-238.

relying on the term ‘mother’.⁶² Careful drafting could minimise, if not eliminate, the risks underlying the Court’s reluctance to initiate reform. Therefore, it has been suggested that the CoA, having acknowledged the significant infringement of the appellants’ rights, should have made a declaration of incompatibility under s 4 HRA 1998. This argument will now be examined, and it will be shown that the Court’s decision was appropriately attuned to the unacceptably high political price to be paid if a declaration were made in this context.

A. What Would a Declaration Under Section 4 Have Achieved?

Those critical of the outcome in *McConnell* have argued that making a declaration of incompatibility ‘would have been one small step towards addressing a much wider issue and could have provided the impetus needed for Parliament to re-evaluate our understanding of legal parenthood to better reflect the complexities of modern family forms’.⁶³ Such views undoubtedly hold some force. The notion that legislatures are ill-suited to interpret the rights of the truly unpopular⁶⁴ is acknowledged. Unfortunately, the political reality is such that the transgender (and the wider LGBTQ+) community still face opposition to the recognition of their very existence, let alone their fundamental rights. This opposition was vocalised loud and clear in a recent debate in the House of Lords, where many members of the House expressed their disdain at the proposed use of gender-

⁶² Fenton-Glynn (n 7) 36.

⁶³ *ibid* 37.

⁶⁴ Kent Roach, ‘The Varied Role of Courts and Legislatures in Rights Protection’ in Murray Hunt and others (eds), *Rights Parliaments and Human Rights* (Hart Publishing 2015) 420.

neutral language in maternity legislation.⁶⁵ Baroness Lady Noakes gave a scathing speech in which she argued that gender inclusive language only appeals to those in ‘woke Brighton’, whilst appalling men and women in ‘mainstream Britain’.⁶⁶ Perhaps the CoA’s suggestion that the transgender community would be able to persuade Parliament to take a different view in time was an optimistic one.

An assessment of some of the most progressive jurisdictions in this context reveals that judicial declarations of incompatibility are often the most effective tool in achieving policy change. Snow explains how LGBTQ+ Canadians have made considerable policy gains in the courtroom, for example with the implementation of the All Families are Equal Act introduced in Ontario in 2016.⁶⁷ The legislation was introduced to ensure equal treatment for all parents and children, and to update the province’s previous legislation which did not explicitly address conception through ART or third-party arrangements. The government also took the opportunity to replace all references to ‘mothers’ and ‘fathers’ on government forms and, where possible, in existing laws, to ‘parents’ and ‘guardians’.⁶⁸ Importantly for this purpose, the success of the Bill which led to the Act’s implementation was partly attributable to a series of judicial declarations of incompatibility. Several judges declared Ontario’s former parentage regime to be incompatible with s 15 of the Canadian Charter of Rights and Freedoms, which

⁶⁵ HL Deb 22 February 2021, vol 810, cols 636-691.

⁶⁶ *ibid* col 640.

⁶⁷ All Families Are Equal Act (Parentage and Related Registrations Statute Law Amendment), 2016 c. 23 (Ontario).

⁶⁸ For example, the definition of s 1 of the Vital Statistics Act, 1990 c. V.4 was amended by striking out ‘from its mother’ and substituting ‘from a person’.

guarantees equality under the law.⁶⁹ Snow emphasises the staying power that such a judicial victory has on equality rights in legislative debate. He asserts that for this reason, Charter litigation is an especially attractive (and effective) strategy for those seeking policy change.⁷⁰ A declaration of incompatibility in *McConnell* could have provided a similar impetus for legislative reform.

B. A Declaration of Incompatibility Was Inappropriate in McConnell

Nonetheless, the CoA's decision to refuse to make a declaration of incompatibility was both legally and politically correct. As the CoA observed, when assessing the compatibility of primary legislation with Convention rights, this has to be performed just as it would be by the Strasbourg Court at the time when the case comes before the Court.⁷¹ At present, there is no European consensus on the issue as to whether transgender men who give birth should be permitted to be registered as the legal 'father' or 'parent' of their child. If anything, the practice of most countries in the Council of Europe is to register those men as 'mothers'.⁷² Therefore, this is almost certainly an area where the ECtHR would find that there is a wide margin of discretion left to Member States. The CoA noted that the German case mentioned in Section 4 of this paper is currently pending before the ECtHR.⁷³ The reasoning in that case was described by the CoA

⁶⁹ Dave Snow, 'Litigating Parentage: Equality Rights, LGBTQ Mobilization and Ontario's *All Families Are Equal Act*' (2017) 32 *Canadian Journal of Law and Society* / *Revue Canadienne Droit Et Société* 329, 334 (see discussion on *Grand v Ontario* (2016)).

⁷⁰ *ibid* 332.

⁷¹ *McConnell* EWCA (n 3) [42].

⁷² *Margaria* (n 8) 243.

⁷³ *McConnell* EWCA (n 3) [73].

as ‘compelling’,⁷⁴ suggesting that it agreed that this matter will likely be deemed to be within the margin of appreciation. In any event, the CoA did not think it was appropriate to pre-empt the ECtHR’s decision.⁷⁵ This reinforces the CoA’s view that, unless and until Strasbourg indicates otherwise, the issues raised in *McConnell* were best left to the expertise of Parliament.

Moreover, the political consequences of a declaration of incompatibility must not be underestimated. The exceptional nature of s 4 has repeatedly been emphasised by the UK courts. In *Ghaidan v Godin-Mendoza*, Lord Steyn made clear that ‘resort to s 4 must always be an exceptional course’.⁷⁶ Judicial reluctance to exercise the power under s 4 is partly due to the consequences of making such a declaration. While in theory, once a declaration is made, it is for the government to decide whether to amend the legislation or not, the political reality is such that this discretion is much narrower than it initially appears. For example, in the Ontarian context, MacFarlane has observed that politicians ‘face a particular symbolic difficulty in being seen as infringing equality’,⁷⁷ such that the political price to pay is not worth even the appearance of violating equality rights.⁷⁸

Snow refers to this phenomenon as ‘rights talk’.⁷⁹ A declaration of incompatibility with Charter rights can effectively neutralise conservative opposition to progressive legislative reform. Of course, this will sometimes be advantageous. It would

⁷⁴ *ibid* [78].

⁷⁵ *ibid* [78].

⁷⁶ *Ghaidan v Godin-Mendoza* [2004] UKHL 30, [2004] 2 AC 557 [50].

⁷⁷ Emmett MacFarlane, ‘The Court in Government and Society: Dialogue, Public Opinion and the Media’ in *Governing from the Bench* (UBC Press 2013, Vancouver) 47.

⁷⁸ Snow (n 69) 332.

⁷⁹ *ibid* 338-341.

enable the kind of reform advocated for in *McConnell* to overcome purely transphobic opposition. Conversely, ‘rights talk’ risks silencing valid counterarguments or suggested qualifications to legislative reform and equips campaigners with a dangerously powerful political sword. For example, at the amendment stage of the All Families are Equal Act, the choice to leave ‘parent’ as the sole marker for parental identification on a child’s birth certificate was challenged. A conservative Ontarian MPP (rather sensibly) suggested an amendment that would have given parents the choice of being identified as a ‘mother’, ‘father’, *or* ‘parent’.⁸⁰ Advocates for this amendment were criticised by liberal MPPs for ‘choosing to align themselves with right-wing, socially conservative groups that do not want LGBTQ2+ families, and families who use assisted reproduction, to be treated equally.’⁸¹ The amendment was ultimately rejected.

If a declaration of incompatibility had been made in *McConnell*, campaigners would have been able to rely on the incompatibility with the ECHR in political discourse. As was the case in Ontario, this additional bargaining power could have eventually resulted in reforms that go further than necessary to record transgender parents in a way that does not contradict their confirmed gender. The importance of the concerns raised by those criticising the case is acknowledged, and the need to protect the rights of the transgender community is not an issue this paper takes lightly. However, the author argues that the Court was right to proceed with caution and to refuse to make a declaration of incompatibility under s 4 HRA 1998. By doing so, it reduced the risk of radical reform being pushed through without meticulous and balanced consideration of its effects.

⁸⁰ *ibid* 342.

⁸¹ Discussed in Snow (n 69) 342.

6. *A Wider Issue*

Despite maintaining that the Court made the correct decision in *McConnell*, it is admitted that the law's heteronormative goal to ensure each child is assigned to its parents in accordance with biological reality is outdated. The real issue at the heart of most criticisms directed towards *McConnell* is that the Court was restricted by outdated definitions that prioritised the heteronormative nuclear family. The refusal to make a declaration of incompatibility does not entail that *McConnell* ought to be perceived as a complete missed opportunity to resolve this problem. The proceedings could still provide the much needed launching pad for increased scrutiny and pressure on the government. They have given this important issue the publicity it deserves—something which might not have been achieved in the near future otherwise. This trajectory should be seized upon, and reform to reflect social reality more accurately should be considered.

A. The Law's Failure to Capture Modern Families

Family law is constantly evolving, and recent developments mean that there is no longer consensus on what it means to be a parent in social or legal terms. The law has undoubtedly taken a step in the right direction in recent years by beginning to recognise socially constructed parenthood through adoption, parental orders, and the provisions of the HFE Act 2008.⁸² Yet, it can still be criticised for the fact that untraditional forms of parenthood

⁸² Thérèse Callus, 'What's the Point of Parenthood? The Agreed Parenthood Provisions Under the HFE Act 2008 and Inconsistency with Intention' (2019) 41 *Journal of Social Welfare and Family Law* 389, 398-399.

are retrofitted into conventional frameworks rather than recognised in their own right. A good example is where a lesbian couple conceives a child through ART. Although the female partner of the woman who gives birth is recognised as a ‘second female parent’ (in certain pre-determined circumstances),⁸³ the law is still unable to accommodate for intentional parental projects involving more than two parents. For instance, where a lesbian couple has a parental project with the biological father of the child, the law is unable to capture the role of a third parent. The result is that one parent must inevitably forego the legal recognition of his or her parental role. The increasing use of ART to realise unconventional family forms means that the time has come for the law to embark upon a thorough review of the law on parenthood to move away from mimicking genetic parentage to encompass wider parenthood relationships.⁸⁴ Such reform would not only benefit transgender persons, but also those other rightsholders considered in the CoA’s judgment, and the LGBTQ+ community more broadly.⁸⁵

B. What Would Reform Look Like?

The GRA itself has been criticised for its failure to capture gender-fluid and non-binary people who do not adhere to traditional gendered terminology. The government recently considered reforming the Act, although debates centred on the accessibility of obtaining a GRC and self-identification, rather than the allocation of legal parenthood.⁸⁶ Perhaps there is an opportunity here that could be seized upon for Parliament to undertake a broader review and examine the issues highlighted in

⁸³ See Human Fertilisation and Embryology Act 2008, ss 44-47.

⁸⁴ Callus (n 82) 400.

⁸⁵ Russel (n 24).

⁸⁶ HC Deb 24 September 2020, vol 680, cols 1131-1139; HL Deb, 25 September 2020), vol 805, cols 2001-2004.

McConnell. Whether such reform should ultimately be considered desirable, and if so, what form it should take is beyond the scope of this paper and, in any event, would require a thorough review and consultation of the law. There are, however, a few important considerations that ought to be taken into account.

First and foremost, any reforms must be well-informed and extensively drafted to ensure that they strike an appropriate balance between the interests of transgender parents, cisgender parents, and their children. The balancing of these interests and the extent of any reform could only be assessed by a democratically elected Parliament, and even then, it will not be an easy task. However, as alluded to in the discussion of the Ontarian reforms, Parliament should be very slow to remove the terms ‘mother’ and ‘father’ entirely. In the 1990s, Fineman warned of the undesirable consequences of de-gendering the family including ignoring the unique role of mothers in childcare laws.⁸⁷

Even some commentators who fiercely criticised *McConnell* and the law’s failure to protect transgender rights have admitted that the radical potential involved in de-gendering legal parenthood might ‘skip a step’ in the ongoing fight for equality.⁸⁸ Margaria argues that replacing mothering with parenting has failed to deliver on its promises and that no concomitant reconstruction of the institution of fatherhood has taken place. A man’s legal status in relation to a child is still primarily dependent upon his relationship with the child’s mother. Where a man is married to a woman who gives birth to a child, the default rule is that he is to be named on the birth certificate as the legal father of that child. This disjunction is particularly stark in surrogacy

⁸⁷ Martha Fineman, *The Neutered Mother, the Sexual Family, and Other Twentieth Century Tragedies* (Routledge 1994) ch 8.

⁸⁸ Margaria (n 8) 246.

cases where the surrogate's husband is likely neither to be related, nor intended to be a primary caregiver of the child.⁸⁹

Of course, surrogacy law is riddled with its own policy concerns. However, the rhetoric that every child *needs* to have a mother (in law) and that a man's legal status can still be determined by virtue of his relationship with the mother is part of a deeper societal issue. Unmarried men are generally regarded as being free to walk away from their caring (i.e., non-financial) responsibilities towards children. De-gendering parenthood will not resolve the underlying discriminatory societal practices that necessitated an elevated status of motherhood in the first place.

It seems that the best path forward would be to retain the term 'mother' to ensure that both the rights of parents who wish to be so identified and other statutes reliant on that term are unaffected. Parliament could introduce the term 'gestational parent' as an opt-in, alternative status to motherhood that would provide identical rights. This has the advantages of retaining the accuracy of the birth registration scheme, whilst providing transgender *or* cisgender parents with a gender-neutral alternative.

Existing schemes could be consulted when considering such reform. For example, the provisions governing the determination of parenthood in the context of ART could prove helpful. Sections 35 to 37 of the HFE Act 2008 outline the provisions for the determination of legal parenthood of the second (i.e., non-gestating) parent for heterosexual couples. Meanwhile, ss 44-47 provide equivalent provisions for lesbian couples. This demonstrates that it is possible to devise concurrent schemes of protection that encompass both 'traditional' and 'unconventional' family forms. Moreover, where the criteria in ss

⁸⁹ See for example: *R (on the application of H) v Secretary of State for Health and Social Care* [2019] EWHC 2095 (Admin).

44-47 are fulfilled, the female partner of the mother is recognised in law as the second ‘female parent’ of the child. The law has thereby already recognised legal parenthood beyond ‘mothers’ and ‘fathers’. There is no obvious policy reason to reject the term ‘gestational parent’ as a legally equivalent alternative to the term ‘mother’.

Under such a proposal, the statutes that currently rely on the term ‘mother’ to allocate legal rights or duties could be amended to acknowledge an equivalent legal status of ‘gestational parents’.

Conclusion

The CoA in *McConnell* delivered a coherent and well-reasoned judgment that was justified in its conclusion. The rights of YY and children generally to ascertain definitively their biological origins were rightly considered to outweigh the individual interests at the centre of the dispute. The Court correctly identified the risks inherent in allowing Mr McConnell’s appeal for other areas of family law and it is suggested that this left the judges with no choice other than to reject the appellants’ proposals. The fight for equality is ongoing. If the transgender community and its supporters are to achieve systemic change, its attention should be redirected towards Parliament. The radical overhaul of the law required to achieve what was being argued for in *McConnell* could only be implemented following a thorough review of the law to assess whether such reform is normatively desirable. If this is answered in the affirmative, Parliament would next need to undertake an extensive consultation to determine the nature of any such reform. What is clear, is the pressing need for

such discussions to take place. As long as the rights and equality of minority groups are being compromised, the law will fail to protect the most vulnerable.

The Future of the Energy Charter Treaty in a post-Achmea Era

Sze Hian Ng*

Abstract—The CJEU’s decision in *Achmea* marked the end of intra-EU Bilateral Investment Treaties, due to their incompatibility with the autonomy of the EU legal order. Subsequent questions have surfaced over whether *Achmea* has the effect of prohibiting disputes between EU Member States under the Energy Charter Treaty (ECT), a Multilateral Investment Treaty of great importance in the field of Investor-State Dispute Settlement. This article assesses the strengths of the legal arguments for limiting or expanding the scope of *Achmea*. While arguments on both sides are compelling, *Achmea* should not have the effect of precluding intra-EU disputes under the ECT in the absence of a CJEU ruling. Moreover, going beyond the legal debate, the changing political climate requires us to examine reforms to the ECT that will preserve the autonomy of EU law. It is suggested that inspiration may be drawn from the Comprehensive Economic and Trade Agreement between EU and Canada, which included an Investment Court System that the ECT can potentially follow.

* The London School of Economics. I am grateful to Ms Natalie Teoh and the OUULJ editorial board for their helpful comments on earlier drafts. Any errors remain my own.

Introduction

On 5 May 2020, 23 EU Member States adopted the Agreement for the Termination of Bilateral Investment Treaties between the Member States of the European Union (the BIT Termination Agreement).¹ In essence, the effect of the BIT Termination Agreement was to terminate all intra-EU bilateral investment treaties (BITs). This landmark moment was the direct result of the CJEU's judgment in *Slovak Republic v Achmea BV (Achmea)*,² where the Grand Chamber held that intra-EU BITs and its Investor-State Dispute Settlement (ISDS) mechanism are incompatible with autonomy of the EU legal order.³ In light of these developments, significant concerns have been raised by various stakeholders that the decision in *Achmea* could be applied to multilateral treaties, in particular, the Energy Charter Treaty (ECT) and the ISDS mechanism enshrined under Art 26 of the treaty.⁴ This would mean that the ECT's ISDS mechanism for

¹ The Agreement for the Termination of Bilateral Investment Treaties between the Member States of the European Union [2020] OJ L169/1.

² Case C-284/16 *Slovak Republic v Achmea BV* [2018] 4 WLR 87.

³ *ibid* [57], [59]; A BIT refers to an investment treaty concluded between Country A and Country B which confers certain substantive rights, such as the obligation for countries not to expropriate foreign investments without due compensation, on investors from both countries. An investor from Country A can bring a claim against Country B, should Country B breach its obligations under the BIT, and vice versa. This process is known as Investor-State Dispute Settlement.

⁴ Communication from the Commission to the European Parliament and the Council: Protection of Intra-EU Investment, at 3-4, COM (2018) 547 final (July 19, 2018); European Commission, 'Declaration of the Member States of 15 January 2019 on the legal consequences of the *Achmea* judgment and on investment protection' (17 January 2019) <https://ec.europa.eu/info/publications/190117-bilateral-investment-treaties_en> accessed 9 May 2021.

intra-EU disputes will be in conflict with EU law, prompting the EU and its Member States to seek an amendment to the ECT's ISDS mechanism to ensure that it is in line with the EU legal framework.⁵

This article will first examine the reasoning of the CJEU in *Achmea* and how it gave EU Member States the impetus to terminate all intra-EU BITs. The potential implications of this on the future of the ECT will then be explored. It will be concluded that the arguments for the invalidation of the ECT's ISDS mechanism with respect to intra-EU disputes should not be favoured in the absence of a ruling from the CJEU. Lastly, regardless of the legal debate, the political trajectory of the EU and its Member States who wish to reform the ECT due to growing frustration with the current ISDS system must be addressed. As such, it is crucial that Contracting Parties modernise the ECT such that it is capable of simultaneously preserving the ECT's ISDS mechanism and avoiding clashes with the autonomy of EU law by upholding the constitutional structure of the EU.

1. *The Decision in Achmea*

The *Achmea* judgment arose out of a dispute between Achmea, a Dutch insurance company, and the Slovak Republic, after

⁵ Markus Beham and Désirée Prantl, 'Intra-EU Investment Reform: What Options for the Energy Charter Treaty?' (*Kluwer Arbitration Blog*, 7 January 2020) <<http://arbitrationblog.kluwerarbitration.com/2020/01/07/intra-eu-investment-reform-what-options-for-the-energy-charter-treaty/>> accessed 6 November 2020.

legislative changes by the Slovakian government prohibited Achmea from distributing its profits earned from offering private sickness insurance policies in Slovakia.⁶ Achmea thus brought a claim before an arbitral tribunal against the Slovak Republic pursuant to Art 8 of the Dutch-Slovak BIT, which provided for an ISDS mechanism. Despite the Slovak Republic challenging the tribunal's jurisdiction, an award of EUR 22.1 million was eventually rendered in favour of Achmea.

As the seat of arbitration was in Germany, the Slovak Republic subsequently sought to set aside the award before the German Federal Court on the grounds that Art 8 of the BIT was incompatible with Arts 267 and 344 TFEU.⁷ Noting that the CJEU had yet to provide an answer for this question, the German Federal Court requested for a preliminary ruling.

The CJEU firstly referred to Art 344 TFEU, under which 'Member States undertake not to submit a dispute concerning the interpretation or application of the Treaties to any method of settlement other than those provided for therein.'⁸ In other words, according to Art 344 TFEU, Member States agree that the only forums which can interpret or apply EU law are those that are part of the judicial system established by the Treaties. This judicial system, which includes national courts and the CJEU, ensures that the autonomy of the EU legal system is preserved by interpreting EU law in a consistent and uniform manner.⁹ As such, an international treaty, such as an intra-EU BIT, cannot have the effect of shifting the allocation of powers fixed by the

⁶ *Achmea* (n 2) [7]-[8].

⁷ *ibid* [9]-[12].

⁸ Art 344 TFEU.

⁹ *Achmea* (n 2) [32]-[36].

Treaties, and allowing other tribunals outside the judicial system of the EU to interpret EU law.

Moreover, the CJEU was keen to emphasise that Member States must cooperate with one another to respect the autonomy of EU law.¹⁰ The importance of autonomy was first examined in detail by the CJEU in the landmark judgment, Opinion 2/13 on the accession of the EU to the European Convention of Human Rights (ECHR).¹¹ In essence, the autonomy of EU law is best described as a ‘body of principles drawn from the specific characteristics arising from the very nature of EU law’.¹² Such qualities of autonomy include EU law being an independent source of law, the primacy of EU law over national laws of the Member States, and its direct application to nationals and Member States themselves.¹³ These aforementioned characteristics of EU law contribute to the constitutional structure of the EU in which ‘mutually interdependent legal relations’ bind the EU and its Member States. It is further observed that the constitutional structure of the EU is upheld and premised on Member States engaging in mutual trust and sincere cooperation to promote a set of common values.¹⁴

The autonomy of EU law is therefore both a structural and existential principle, contributing to the functioning and the very creation of the EU legal order, governing both the internal

¹⁰ Art 4(3) TEU.

¹¹ Opinion 2/13 (Accession of the EU to the ECHR) of 18 December 2014, EU:C:2014:2454; Case C-26/62 *Van Gend en Loos* [1963] CMLR 105 and case C-6/64 *Costa v ENEL* [1964] CMLR 425 which lay down the principle of autonomy of the EU legal system.

¹² Niamh Shuibhne, ‘What is the Autonomy of EU Law, and Why Does that Matter?’ (2019) *Nordic Journal of International Law* 9, 19.

¹³ *Achmea* (n 2) [32]-[36].

¹⁴ Art 2 TEU; see also Opinion 2/13 (n 11) [168].

and external relations of the EU.¹⁵ For example, in its external conception, the principle was employed by the CJEU to determine that EU fundamental rights standards should be prioritised over the content of fundamental rights from external sources such as the ECHR, allowing it to find that the EU could not accede to the ECHR.¹⁶ It is the manner in which the principle of autonomy regulates such external relationships between the EU and other actors in the international plane that is of relevance to our present inquiry.

In this context, the CJEU in *Achmea* noted that Art 267 TFEU, which provides for the preliminary ruling procedure, is the jewel in the crown of the judicial system¹⁷ as it serves to secure the consistency and autonomy of EU law.¹⁸ Through preliminary rulings, the CJEU is able ensure that EU law is applied uniformly by national courts. Indeed, without the ability of national courts to refer questions to the CJEU, divergences between courts in the Member States as to the interpretation of EU law would be liable to jeopardise the very unity of the EU's legal order, and hence its autonomy.¹⁹

¹⁵ Jed Odermatt, 'When a Fence Becomes a Cage: The Principle of Autonomy in EU External Relations Law' (2016) *EUI MWP* 2016/07, 5.

¹⁶ *Shuibhne* (n 12) 21.

¹⁷ *Achmea* (n 2) [37]; see also Paul Craig, 'The Jurisdiction of Community Courts Reconsidered' (2001) 36 *Texas International Law Journal* 555, 559.

¹⁸ Marie Stoyanov and Lucia Raimanova, 'The Court of Justice of the European Union finds the arbitration provision in The Netherlands-Slovakia BIT incompatible with EU law' (*Allen & Overy*, 30 March 2018). <<https://www.allenoverly.com/en-gb/global/news-and-insights/publications/arbitration-provision-in-the-netherlands-slovakia-bit-incompatible-with-eu-law>> accessed 7 November 2020.

¹⁹ Case C-314/85 *Foto-Frost* [1988] 3 *CMLR* 57 [15].

This concern was the very reason why the CJEU found that Art 8 of the Dutch-Slovak BIT was incompatible with EU law. Examining Art 8 closely, the CJEU observed that Art 8(6) of the Dutch-Slovak BIT requires the arbitral tribunal to ‘take account in particular of the law in force of the contracting party concerned’. As Basedow describes, since EU law is intertwined with the national law of EU member states, the arbitral tribunal would thus have to interpret EU law if relevant to a given case.²⁰

This would be in breach of EU law as the arbitral tribunal, which was ad hoc, is not part of the judicial system of the Netherlands or Slovakia, and therefore cannot be classified as a court or tribunal ‘of a Member State’ within the meaning of Art 267 TFEU.²¹ As such, the arbitral tribunal is unable to make a preliminary reference request to the CJEU in relation to the valid interpretation of EU law, which threatens the autonomy and consistency of EU legal order as explained earlier. Given such a tribunal is outside the judicial system of the EU, it also undermines the principle of sincere cooperation should a tribunal interpret EU law.

Furthermore, whilst the CJEU did not explain its reasoning for this, it held that the existence of such an arbitral tribunal also contravened Art 344 TFEU. One possible rationale is that by submitting a dispute that may potentially involve the interpretation of EU law to an ad hoc arbitral tribunal outside the judicial system of the EU, the allocation of powers fixed by the Treaties is shifted, thereby breaching Art 344 TFEU.

²⁰ Robert Basedow, ‘The *Achmea* Judgment and the Applicability of the Energy Charter Treaty in Intra-EU Investment Arbitration’ (2020) 23 *Journal of International Economic Law* 271.

²¹ *Achmea* (n 2) [45]-[46].

In summary, the CJEU decided that Arts 267 and 344 TFEU must be interpreted as precluding ISDS provisions in BITs between EU Member States.²² Central to the CJEU's reasoning is the idea that intra-EU BITs infringe on the autonomy of the EU legal order, which requires EU law to be interpreted in a uniform and consistent manner.²³

2. *The BIT Termination Agreement and the ECT*

The decision in *Achmea* was widely accepted as signalling the end of intra-EU BITs. On 15 January 2019, a majority of the EU Member States released a declaration, expressing their intention to terminate all 190 intra-EU BITs, given that the ISDS arbitration clauses in these treaties are contrary to EU law and therefore inapplicable.²⁴ This was swiftly followed by the adoption of the BIT Termination Agreement on 5 May 2020,

²² *ibid* [60].

²³ *ibid* [35].

²⁴ European Commission, 'Declaration of the Member States of 15 January 2019 on the legal consequences of the *Achmea* judgment and on investment protection' (17 January 2019) <https://ec.europa.eu/info/publications/190117-bilateral-investment-treaties_en> accessed 9 May 2021; see also Nikos Lavranos, 'The EU Plurilateral Draft Termination Agreement for All Intra-EU BITs: An End of the Post-Achmea Saga and the Beginning of a New One' (*Kluwer Arbitration Blog*, 1 December 2019) <<http://arbitrationblog.kluwerarbitration.com/2019/12/01/the-eu-plurilateral-draft-termination-agreement-for-all-intra-eu-bits-an-end-of-the-post-achmea-saga-and-the-beginning-of-a-new-one/>> accessed 7 November 2020.

which formalised the termination of all intra-EU BITs, and deprived any sunset clauses of their legal effects.²⁵

Nonetheless, the BIT Termination Agreement stopped short of covering Intra-EU ISDS grounded in the Energy Charter Treaty.²⁶ In the preamble, it was acknowledged that the agreement does not cover intra-EU proceedings on the basis of Art 26 of the Energy Charter Treaty. It was decided that the matter would be dealt with at a later stage.²⁷

3. Implications of Achmea on Intra-EU Disputes Grounded on the ECT

Known as the ‘child’ of the European Union,²⁸ the ECT is a multilateral treaty that provides a global legal framework which regulates and promotes all forms of energy cooperation, such as trade, investment, and transit.²⁹ Currently, the ECT has 57

²⁵ BIT Termination Agreement (n 1).

²⁶ Matteo Fermeglia and Alessandra Mistura, ‘Killing all birds with one stone: Is this the end of Intra-EU BITs (as we know them)’ (*EJIL: Talk!*, 26 May 2020) < <https://www.ejiltalk.org/killing-all-birds-with-one-stone-is-this-the-end-of-intra-eu-bits-as-we-know-them/> > accessed 7 November 2020.

²⁷ BIT Termination Agreement (n 1).

²⁸ Jan Kleinheisterkamp, ‘The Next 10 Year ECT Investment Arbitration: A Vision for the Future – From a European Law Perspective’ (2011) LSE Legal Studies Working Paper No 7/2011.

²⁹ Ernesto Bonafé and Aurore Vanhay, *The role of the Energy Charter Treaty in fostering regional electricity market integration: Lessons learn from the EU and implications for Northeast Asia* (Energy Charter Secretariat, 2015) 22.

Contracting Parties, including a majority of the EU Member States, and the EU itself.³⁰

In order to strengthen and safeguard the effectiveness of its substantive provisions, Art 26 of the ECT enables investors of a Contracting Party to bring claims against another Contracting Party for a breach of the latter's obligations under the ECT. Ad hoc tribunals and tribunals established under the International Centre for Settlement of Investment Disputes or the Arbitration Institute of the Stockholm Chamber of Commerce have jurisdiction to hear these claims under Art 26(4)(a) ECT. Similar to intra-EU BITs, Art 26 ECT would therefore also allow for investors from one EU Member State to sue another EU Member State, provided they are all Contracting Parties to the ECT.³¹ From this perspective, if the reasoning in *Achmea* is extended to multilateral treaties such as the ECT, intra-EU disputes under the ECT may be incompatible with the autonomy of the EU legal order. It is estimated that cases of intra-EU disputes grounded on Art 26 account for 10% of all known ISDS proceedings worldwide, stressing the importance of this matter in the field of investment-treaty arbitration.³²

³⁰ Signatories/Contracting Parties (Energy Charter Treaty, 18 February 2019) <<https://www.energycharter.org/process/energy-charter-treaty-1994/energy-charter-treaty/signatories-contracting-parties/>> accessed 9 November 2020.

³¹ Art 26 ECT; see also Thomas W Walde, 'European Energy Charter Conference: Final Act, Energy Charter Treaty, Decisions and Energy Charter Protocol on Energy Efficiency and Related Environmental Aspects' (1995) 34 *International Legal Materials* 360, 360.

³² Ciaran Cross and Vivian Kube, 'Is the arbitration clause of the Energy Charter Treaty compatible with EU law in its application between EU Member States' (2018) *The Munich Environmental Institute* 1, 10.

Various stakeholders have expressed vastly differing opinions with regards to the implication of *Achmea* for the future of the ECT. While the interpretation of *Achmea* and its relationship with the ECT will change depending on whether one adopts a public international law perspective, or views EU law as a superior legal order,³³ this article adopts the view that the intra-EU disputes should not be exempted from the purview of Art 26 ECT, especially without a decisive ruling from the CJEU.

A. European Commission and Majority of EU Member States: Intra-EU Disputes under Art 26 ECT are Incompatible with EU law

The European Commission (Commission),³⁴ and a majority of the EU Member States,³⁵ have declared that similar to *Achmea*, intra-EU arbitration under the ECT is incompatible with the autonomy of EU law. Notably, the two parties take slightly different approaches in coming to the same conclusion, with the EU Member States' approach being the more principled of the two.

For the Commission, the reasoning in *Achmea* applies directly to Art 26 ECT. It must be noted that the Commission's rationale for doing so is vague and requires us to make several inferences. The Commission firstly emphasised the primacy of

³³ April Lacson, 'What Happens Now? The Future of Intra-EU Investor-State Dispute Settlement under the Energy Charter Treaty' (2019) 51 NYUJ Intl L & Pol 1327, 1340.

³⁴ Communication from the Commission to the European Parliament and the Council: Protection of Intra-EU Investment, at 3-4, COM (2018) 547 final (July 19, 2018).

³⁵ European Commission, 'Declaration of the Member States of 15 January 2019 on the legal consequences of the *Achmea* judgment and on investment protection' (n 24).

EU law, thus alluding to notions of the EU as an autonomous legal order.³⁶ Presumably, just like the arbitral tribunal in *Achmea*, an arbitral tribunal constituted pursuant to Art 26 ECT may be required to interpret EU law. This is reasonable: as Kleinheisterkamp argues, EU law and the ECT overlap notably in relation to the protection and enforcement of investor rights.³⁷ As such, Art 26 ECT, analogous to Art 8 in the Dutch-Slovak BIT, would open the possibility of investors submitting disputes to a tribunal outside of the EU's judicial system. Here, said tribunal may be called upon to interpret EU law provisions, *without* the capacity to make a preliminary reference to the CJEU. This effectively contravenes Arts 267 and 344 TFEU as discussed earlier, and the incompatibility with the autonomous nature of the EU legal order means that intra-EU disputes under Art 26 ECT should be precluded.

Moreover, the Commission argues that the EU being a Contracting Party to the ECT does not affect the above findings. To them, the participation of the EU in the ECT 'has only created rights and obligations between the EU and third countries and has not affected the relations between the EU Member States'.³⁸ There was, however, no further justification for this distinction, making such an assertion questionable. According to Basedow, given that the EU has ratified the ECT as an international agreement, it forms part of EU law. Based on Art 3(5) TEU, the EU has to abide by a 'strict observance' of international law,³⁹ which suggests that the ECT should at least in principle prevail

³⁶ Communication from the Commission (n 34).

³⁷ Jan Kleinheisterkamp, 'Investment Protection and EU Law: The Intra- and Extra-EU Dimension of the Energy Charter Treaty' (2012) 1 *Journal of International Economic Law* 15, 99.

³⁸ Communication from the Commission (n 34) 4.

³⁹ Art 3(5) TEU.

over EU law where a conflict exists.⁴⁰ This argument appears fatal to the Commission's opinion.

Nonetheless, the exact status of international treaties in EU law is not as clear-cut as Basedow contends. Ziegler argues that international law, once directly applicable within the EU legal order, does not typically rank more highly than EU primary law.⁴¹ In other words, from this perspective, the ECT has become part of the EU legal order, and affects both the relations between EU Member States and between the EU and third countries. However, it remains the case that the ECT must be compatible with EU law. On this basis, intra-EU disputes under Art 26 ECT must be precluded in light of *Achmea*. There are thus some grounds for supporting the Commission's view, although it is incorrect in its assertion that the ECT does not affect the relations between EU Member States.

In contrast, the reasoning behind the stance adopted by a majority of the EU Member States corrects the mistake made by the Commission. It accepts that international agreements concluded by the EU, including the ECT, are 'an integral part of the EU legal order, and must therefore be compatible with the Treaties'.⁴² Following from this premise, the same reasoning in *Achmea* can be applied: an arbitral tribunal established under Art 26 ECT will be in violation of Arts 267 and 344 TFEU and will

⁴⁰ Basedow (n 20) 275.

⁴¹ Katja Ziegler, 'The Relationship between EU Law and International Law' University of Leicester School of Law Research Paper No 15-04, 11.

⁴² European Commission, 'Declaration of the Member States of 15 January 2019 on the legal consequences of the *Achmea* judgment and on investment protection' (n 24), also see Case C-266/16 *Western Sahara* [2018] 3 CMLR 15, [42]-[51] where the CJEU holds that international agreements ratified by the EU form part of the EU legal order, and must be compatible with the Treaties.

consequently be incompatible with the Treaties. A tribunal, which is not part of the judicial system of the EU and is unable to make preliminary references to the CJEU, cannot be allowed to interpret EU law. To do so would be to undermine EU law as an independent, autonomous source of law, infringing on the principle of mutual trust between Member States and ultimately subverting the constitutional structure of the EU.

B. Minority of EU Member States and Arbitral Tribunals: Achmea Has No Implication for Intra-EU Disputes

Yet, for a host of reasons, other stakeholders—including EU Member States, arbitral tribunals, national courts, and academic commentators—remain unconvinced by the Commission and Member States' efforts to carve out intra-EU disputes from the jurisdiction of Art 26 ECT.

For one, a minority of EU Member States, including Finland, Slovenia, and Luxembourg, noted that the *Achmea* judgment made no reference to Art 26 ECT whatsoever.⁴³ In *Achmea*, the CJEU deliberately distinguished intra-EU BITs from international agreements which were concluded by the EU itself, which is why the three states argued that its ruling can only be

⁴³ Press Release, European Commission, Declaration of the Representatives of the Government of the Member States, of 16 January on the Enforcement of the Judgment of the Court of Justice in *Achmea* and on Investment Protection in the European Union 3 (Jan 16, 2019), <<https://www.regeringen.se/48ee19/contentassets/d759689c0c804a9ea7af6b2de7320128/achmea-declaration.pdf>> accessed 14 May 2021.

applied to other intra-EU BITs similar to that of the Dutch-Slovak BIT.⁴⁴

It is possible the CJEU was simply erring on the side of caution, given that it was not asked by the German Federal Court to determine whether Art 26 ECT was incompatible with EU law. The CJEU did ultimately acknowledge that an international agreement ratified by the EU and providing for the establishment of a court responsible for the interpretation of its provisions must still respect the autonomy of the EU and its legal order.⁴⁵ This seems to be in line with the aforementioned notion that international agreements which form part of the EU legal order must still be compatible with EU primary law.⁴⁶

However, the minority group of EU Member States also made an important reference to a number of international arbitration decisions made post-*Achmea* which have held that intra-EU disputes under Art 26 ECT are still applicable, and not contrary to EU law.⁴⁷ In particular, the decisions of *Masdar v Spain* (*Masdar*),⁴⁸ *Vattenfall AB v Germany*⁴⁹ (*Vattenfall*) and *Landesbank Baden-Württemberg v Spain*⁵⁰ (*LBBW*) are highly relevant.

Masdar was the first case to decide whether *Achmea* had any bearing on intra-EU disputes based on Art 26 ECT. As discussed above, the tribunal similarly noted that the scope of the

⁴⁴ *Achmea* (n 2) [58].

⁴⁵ *ibid* [57].

⁴⁶ Ziegler (n 41).

⁴⁷ Press Release, Minority of EU Member States (n 43).

⁴⁸ *Masdar Solar & Wind Cooperatief U.A. v Kingdom of Spain*, ICSID Case No. ARB/14/1, Award.

⁴⁹ *Vattenfall AB, et al. v Federal Republic of Germany*, ICSID Case No. ARB/12/12, Decision on the *Achmea* Issue.

⁵⁰ *Landesbank Baden-Württemberg et al. v Kingdom of Spain*, ICSID Case No. ARB 15/45, Decision on the 'Intra-EU' Jurisdictional Objection.

decision in *Achmea* was limited to intra-EU BITs, and does not apply to multilateral treaties such as the ECT.⁵¹ It drew support from the Opinion of Advocate General Wathelet,⁵² who acknowledged that *Achmea* was simply an opportunity for the CJEU to express its views on the question of compatibility between *intra-EU BITs* and their ISDS mechanisms with EU law. The Advocate General differentiated the ECT as a multilateral treaty, in which all Contracting Parties, including the EU, participate on an ‘equal footing’. ISDS under Art 26 ECT could operate between EU Member States, and there was no suggestion that it was incompatible with the EU legal order.⁵³ As such, *Achmea* cannot be arbitrarily extended to multilateral treaties such as the ECT. Furthermore, the question of whether the ECT could be reconciled with EU law was not discussed by the CJEU, which led to the tribunal’s conclusion that there is nothing to suggest that the ruling in *Achmea* implies that an EU investor is barred from bringing an investment arbitration proceeding against an EU Member State under Art 26 ECT.

The merits of the tribunal’s findings in *Masdar* lies in its well-reasoned analysis of the scope of *Achmea*’s applicability based on a solid, textual interpretation of the CJEU’s decision. Nonetheless, it does not directly address the arguments made by the Commission and the majority of EU Member States for extending the rationale and reasoning of the CJEU in *Achmea* to preclude intra-EU BITs pursuant to Art 26 ECT. Perhaps for this reason, *Masdar* was quickly followed by a number of tribunals who sought to expand on the reasons why *Achmea* was of no relevance to intra-EU investor-state arbitration proceedings

⁵¹ *Masdar* (n 48) [679].

⁵² Case C-284/16 *Slovak Republic v Achmea BV* [2018] 4 WLR 87, Opinion of AG Wathelet, [23].

⁵³ *ibid* [43].

grounded on Art 26 ECT. In *Vattenfall*, the tribunal produced a detailed 72-page procedural decision examining the complex relationship between EU law and investment treaties as a creature of public international law.⁵⁴ Based on the general principle of treaty interpretation in public international law, the tribunal ultimately concluded that it had the jurisdiction to resolve the dispute between Vattenfall, a Swedish energy company (as Claimant), and Germany (as Respondent) over the latter's alleged breaches of its obligations under the ECT. The decision was centred on the argument that EU law, and the judgments of the CJEU, have no effect on an arbitral tribunal's jurisdiction founded solely on Art 26 ECT.

The tribunal first established that its 'competence to decide the dispute is derived from the consent of the Parties to arbitrate pursuant to the ECT'.⁵⁵ In the absence of any explicit choice of law clause for the law governing the tribunal's jurisdiction, the starting point for its jurisdictional analysis must therefore be Art 26 ECT itself.⁵⁶ Art 26 ECT, like all provisions in international treaties, must be interpreted in accordance with the Vienna Convention on the Law of Treaties (VCLT).

The question is therefore whether EU law, and the *Achmea* decision, can inform the interpretation of Art 26 ECT through the provisions under the VCLT.⁵⁷ This was possible

⁵⁴ Kirstin Schwedt and Hannes Ingwersen, 'Intra-EU ECT Claims Post-Achmea: Vattenfall Decision Paves the Way' (*Kluwer Arbitration Blog*, 13 December 2018) <<http://arbitrationblog.kluwerarbitration.com/2018/12/13/intra-eu-ect-claims-post-achmea-vattenfall-decision-paves-the-way/>> accessed 14 November 2020.

⁵⁵ *Vattenfall* (n 49) [124].

⁵⁶ *ibid*, see also Schwedt and Ingwersen (n 54).

⁵⁷ The tribunal had rejected an alternative submission from Respondent that EU law and the *Achmea* judgment applied under Art

based on Art 31(3)(c) VCLT, which provides that when interpreting treaties, ‘any relevant rules of international law applicable in the relations between the parties’ must be taken into account. As such, if EU law can be proven as a source of international law, there may be an argument for the tribunal to consider the *Achmea* judgment in its interpretation of Art 26 ECT, by virtue of Art 31(3)(c) ECT.⁵⁸ While the tribunal accepted that EU law is international law ‘because it is rooted in international treaties’,⁵⁹ it nevertheless rejected that EU Treaties and the *Achmea* decision should be relied upon when construing the meaning of Art 26 ECT.⁶⁰

According to the tribunal, Art 31(3)(c) is ‘not the starting point of the interpretation exercise under the VCLT’.⁶¹ Instead, the general rule of interpretation is enshrined under Art 31(1) VCLT. This requires the tribunal to interpret Art 26 ECT ‘in good faith in accordance with the *ordinary meaning* to be given to the terms of the treaty in their context and in light of its object and purpose.’⁶² Priority is thus given to the text of the ECT itself. However, applying the *Achmea* decision would have the effect of directly contradicting the plain meaning of Art 26 ECT, which is to allow for disputes between an investor of a Contracting Party

26(6) ECT. Art 26(6) ECT provides that the tribunal shall decide the issues in dispute in accordance with the ECT, and applicable rules of principles of international law. Arguably, if EU law was a part of international law, the *Achmea* judgment may have applied through Art 26(6). However, this was rejected by the tribunal on the basis that ‘issues in dispute’ only referred to the merits of the disputes and does not relate to the Tribunal’s jurisdiction; see *Vattenfall* (n 49) [113]-[122].

⁵⁸ *Vattenfall* (n 49) [130]-[135].

⁵⁹ *Electrabel SA v Hungary*, ICSID Case No. ARB/07/19, Award, 25 November 2015; *Vattenfall* (n 49) [146].

⁶⁰ *Vattenfall* (n 49) [152].

⁶¹ *ibid* [153].

⁶² Art 31(1) VCLT (emphasis added).

and a Contracting Party to be resolved by way of arbitration. From this perspective, it is illogical to interpret the treaty such as to preclude intra-EU disputes,⁶³ especially since Art 26 ECT did not explicitly provide for such a carve-out. Indeed, it has been noted by a number of tribunals that the EU should have included a ‘disconnection clause’ that shields intra-EU relationships from the scope of the Art 26 ECT if that was its intention.⁶⁴ The argument is even more persuasive when one undertakes an examination of the *travaux préparatoires*, which reveal that the EU could not have intended for Art 26 ECT to preclude intra-EU BITs.⁶⁵ According to Basedow, while the EU initially pressed for a disconnection clause, it subsequently conceded this demand due to hesitation from the other negotiating parties, therefore accepting the risks of Art 26 ECT’s intra-EU applicability.⁶⁶

Moreover, the tribunal considered that allowing *Achmea* to inform the interpretation of Art 26 ECT would produce an ‘incoherent and anomalous result’ as it creates multiple interpretations of the same provision.⁶⁷ This is in contrary to the requirement of terms of treaties having a single consistent meaning, in line with the *pacta sunt servanda* principle (Art 26 VCLT). By carving out intra-EU disputes from Art 26 ECT, the meaning and effect of Art 26 ECT will change according to the parties involved in the dispute.⁶⁸ A tribunal would have jurisdiction to adjudicate a dispute between an EU Member State and a non-EU Member State but would lack jurisdiction to do so

⁶³ *Vattenfall* (n 49) [154].

⁶⁴ *Charanne BV & Investments S.A.R.L. v Spain*, SCC Case No. 062/2012, Award [433]-[436]; *Masdar* (n 48) [310]-[314]; *Vattenfall* (n 49) [182]; Basedow (n 20) 275.

⁶⁵ Basedow (n 20) 287-291.

⁶⁶ *ibid.*

⁶⁷ *Vattenfall* (n 49) [155]-[157].

⁶⁸ *ibid.*

if the dispute is between two EU Member States. Such an interpretation is prejudicial to legal certainty, and should not be endorsed unless explicitly provided for.

Notwithstanding the above arguments, the tribunal also held that Art 16 ECT could provide for a ‘simpler and clearer’ route to the same conclusion.⁶⁹ Art 16 ECT is a conflict of laws rule, which states that prior and subsequent international agreements between Contracting Parties cannot have the effect of derogating from the Dispute Settlement provisions of the ECT, where such provisions are more favourable to the Investor. A clash between Arts 267 and 344 TFEU and Art 26 ECT must thus be decided in favour of the latter, since Art 26 ECT is advantageous to the investor, who is empowered to bring a claim against Contracting Parties according to the investor-friendly provisions of the ECT.⁷⁰

As observed by Schwedt and Ingwersen, the ‘carefully crafted’ decision in *Vattenfall* may serve as a blueprint for tribunals in similar cases, with a similar line of reasoning being adopted by the tribunal in *LBBW*.⁷¹ The reasoning in *Vattenfall* is robust, and poses a direct challenge to the views held by the Commission and the majority of the EU Member States by outlining how, as a matter of public international law, the interpretation of Art 26 ECT cannot be defeated by the primacy and autonomy of the EU legal order.

⁶⁹ *ibid* [192].

⁷⁰ Mathias Baudena, ‘Investor-State Dispute Settlement: Understanding the System’s Legitimacy Crisis in Constitutional Terms’ (*LSE Law Review Blog*, 18 February 2021) <<https://blog.lselawreview.com/2021/02/investor-state-dispute-settlement/>> accessed 17 March 2021.

⁷¹ Schwedt and Ingwersen (n 54).

The two opposing arguments and their differences in reasonings boils down to a difference in perspective.⁷² The CJEU is an EU institution, and therefore views the EU legal order as an autonomous, superior legal order. Any international agreement must comply with the EU Treaties, which are accorded the status of primary law. It will therefore hesitate to allow for any derogation from Arts 267 and 344 TFEU. Conversely, arbitral tribunals approach the matter from a public international law perspective. The crux of the issue lies in the interpretation of Art 26 ECT and whether it confers jurisdiction on the tribunal. Since Art 26 ECT is a creature of international law, it is interpreted based on principles of international law, which gives the ECT and the TFEU equal weight as international treaties.

In the face of compelling arguments made by both sides, there is no clear-cut answer to the implications of *Achmea* for Art 26 ECT. Nonetheless, it is suggested that the views adopted by the various tribunals and a minority of the Member States are preferred. The Commission's position relies heavily on extending the rationale of *Achmea* to multilateral treaties, but fails to acknowledge that the CJEU made clear that its decision can only be applied to BITs such as the Dutch-Slovak BIT it was asked to examine. Even if a more generous, contextual reading of *Achmea* is applied, the EU law perspective should not be favoured over orthodox public international law. Arbitral tribunals established under the ambit of Art 26 ECT determine their competence to hear claims based on Art 26 itself and in accordance with the principles of interpretation laid down in the VCLT. Consequently, while Art 26 ECT is likely incompatible with the autonomy of the EU legal order since tribunals may be required to interpret provisions of EU law, the outcome of barring intra-

⁷² Lacson (n 33) 1340.

EU disputes is equally incompatible with the ordinary meaning of Art 26 ECT, the *travaux préparatoires*, and Art 16 ECT.

4. Reforms and Looking Ahead

In view of the described deadlock and the lack of a decision from the CJEU on this specific issue, intra-EU disputes should arguably not be precluded from Art 26 ECT without a decisive and clear CJEU ruling. Simply ignoring the strong arguments for not precluding them would, as noted, be unconvincing and rather requires careful evaluation and reasoning. It remains to be seen whether the CJEU will have an opportunity to do so in the near future. As recently as 27 May 2020, the Svea Court of Appeal rejected the Kingdom of Spain's request for a preliminary ruling from the CJEU in order to clarify whether Art 26 ECT is applicable between EU Member States.⁷³ Yet in the current political climate, the desire of the Commission and a majority of the EU Member States to carve out intra-EU disputes from Art 26 ECT cannot be ignored. The EU is growing increasingly frustrated with the field of ISDS, owing to issues surrounding the transparency and accountability of the current ISDS mechanism.⁷⁴ This has also been reflected in its efforts at reforming the ISDS system, from including an Investment Court

⁷³ *Novenergia II—Energy & Environment (SCA) (Grand Duchy of Luxembourg), SICAR v The Kingdom of Spain*, SCC Case No 2015/06, Decision of the Svea Court of Appeal, 27 May 2020.

⁷⁴ Cecilia Malmström, 'Commission proposes new Investment Court System for TTIP and other EU trade and investment negotiations' (*European Commission*, 16 September 2015) <<http://trade.ec.europa.eu/doclib/press/index.cfm?id=1364>> accessed 14 November 2020.

System in the Comprehensive Economic and Trade Agreement (CETA) to participating in the UNCITRAL Working Group III's Investor-State Dispute Settlement Reform. There is therefore a need to take the views of the Commission seriously, and examine possible reform options that will respect the autonomy of the EU legal order.

Several commentators have suggested that it is possible to amend the provisions of the ECT to provide for an explicit carve-out of intra-EU disputes from the existing provisions on dispute settlement in ECT.⁷⁵ This can be achieved through an *inter se* agreement pursuant to Art 41 VCLT, which allows a select group of Contracting Parties to conclude an agreement to modify the treaty between themselves. Alternatively, the EU may choose to rely on Art 42 ECT to pass an amendment to the ECT once the proposal has received approval from three-fourths of the Contracting Parties.

These options are politically and legally difficult. According to Art 41(b) VCLT, *inter se* agreements must not be incompatible with the 'object and purpose' of the treaty, and the modification cannot be prohibited by the treaty itself. Art 2 ECT provides that the purpose of the ECT is to establish a legal framework to promote long-term cooperation in the energy field. As such, the minority of EU Member States who are against the exclusion of intra-EU BITs may object to the modification on the basis that it deprives their investors of this legal framework to enforce their substantive rights under the ECT. Arguably, Art 16 ECT also becomes an obstacle to the *inter se* agreement, since it specifically bars any derogation from the Dispute Settlement provisions of the ECT, where such provisions are more favourable to the Investor. With regards to the amendment under

⁷⁵ Beham and Prantl (n 5); Basedow (n 20).

Art 42 ECT, the high threshold of three-quarters of the Contracting Parties may not be attainable. The EU, Euratom and 22 of its Member States currently form less than half of the total of 57 Contracting Parties.

Yet, there is hope that a satisfactory compromise can be reached through another means. In 2018, the Energy Charter Conference approved a list of topics for the discussion on the modernisation of the Energy Charter Treaty.⁷⁶ In response to this, the Council of the European Union published its negotiating directives, stating that the EU will strive to ensure that ‘ongoing multilateral reforms of ISDS, such as those within UNCITRAL Working Group III and ICSID, will be applied to the ECT ... including [the provision] of a future Multilateral Investment Court’.⁷⁷ Reference was also made to the need to bring the ECT provisions on ISDS in line with the ‘modern standards of recently concluded agreements by the EU and its Member States’, such that it is in line with the EU legal framework.⁷⁸ The EU’s position was reaffirmed in the Energy Charter Conference’s ‘Policy Options for the Modernisation of the ECT’ on 6 October 2019.⁷⁹

⁷⁶ Energy Charter Secretariat, ‘Approved topics for the modernization for the Energy Charter Treaty’ (International Energy Charter, 29 November 2018)

<<https://www.energycharter.org/media/news/article/approved-topics-for-the-modernisation-of-the-energy-charter-treaty/>> accessed 15 November 2020.

⁷⁷ General Secretariat of the Council, ‘Negotiating Directives for the Modernisation of the Energy Charter Treaty – Adoption’ (Council of the European Union, 2 July 2019) <

<https://data.consilium.europa.eu/doc/document/ST-10745-2019-ADD-1/en/pdf>> accessed 15 November 2020.

⁷⁸ *ibid.*

⁷⁹ Energy Charter Secretariat, ‘Adoption by Correspondence – Policy Options for Modernisation of the ECT’ (Energy Charter conference, Brussels, 6 October 2019)

Given that the discussions for reform within UNCITRAL Working Group III are still ongoing and involve significantly more stakeholders, a quicker solution may therefore lie in the introduction of an ECT Investment Court, modelled after a similar framework under CETA, which was recently deemed as compatible with the autonomy of EU law by the CJEU.⁸⁰ The ECT Investment Court would be characterised by several features that make it distinct from the current ISDS mechanism under Art 26 ECT. Similar to the CETA Investment Court, the ECT Investment Court would only have the jurisdiction to interpret and apply provisions under the ECT, and cannot determine the legality of a measure, alleged to constitute a breach of the ECT, under the domestic law of a Contracting Party.⁸¹ To the extent that it must consider the domestic law of a Contracting Party, it must only do so as a matter of fact, following the prevailing interpretation given to the domestic law by the Courts.⁸² Presumably, this would be akin to the manner in which the English courts deal with foreign law; as a ‘matter of fact that must be proved to the satisfaction of the judge by expert evidence’.⁸³ Such an approach ensures that the autonomy of the EU legal order will not be infringed upon, since the ECT Investment Court will not be required to interpret EU law and its provisions.⁸⁴ There is no need for preliminary references, and the

<<https://www.energycharter.org/fileadmin/DocumentsMedia/CCD/ECS/2019/CCDEC201908.pdf>> accessed 15 November 2020.

⁸⁰ Opinion 1/17 of the Court, EU-Canada CET Agreement, EU:C:2019:341.

⁸¹ *ibid* [120]-[161].

⁸² Art 8.31.2 CETA.

⁸³ *The Kyrgyz Republic v Stans Energy Corporation and Kutisay Mining LLC* [2017] EWHC 2539 (Comm) [18] (Simon Bryan QC).

⁸⁴ Opinion 1/17 (n 80) [131].

allocation of powers fixed by the EU treaties would not be affected.

Conclusion

The decision of the CJEU in *Achmea* has had remarkable implications on the field of international investment law. Apart from the termination of intra-EU BITs, questions still remain over its compatibility with intra-EU disputes grounded on Art 26 ECT. Various stakeholders have thus sought to either restrict or expand the effect of *Achmea*, depending on whether they adopt a public international law perspective, or view EU law as a superior legal order. While both lines of arguments are convincing, it is submitted that intra-EU disputes should not be precluded from Art 26 ECT in the absence of a clear ruling from the CJEU. Beyond the legal debates, it is clear that there is a strong desire on the part of the EU and a majority of its Member States to ensure that the ECT is in line with the autonomy of the EU legal order. Such political trajectory cannot be disregarded, and therefore an adequate reform is required in order to preserve the ECT's ISDS mechanism. This article has proposed that a possible solution may come in the form of an ECT Investment Court. Alternatively, it may be prudent for the Energy Charter Conference to adopt a 'wait-and-see' approach in view of the ongoing reform discussions conducted by the UNCITRAL Working Group III, and seek to apply similar reforms to the ECT.

‘Gentlemen at home, hoodlums elsewhere’: The Extraterritorial Application of the European Convention on Human Rights

Petra Stojnić*

Abstract—This article explores the extraterritorial application of the European Convention on Human Rights, focussing on the implications of the seminal case of *Al-Skeini and Others v United Kingdom*, which explicitly recognised the personal model of jurisdiction, accepted that Convention rights can be divided and tailored and abandoned the concept of *espace juridique*. It concludes that some uncertainties have remained following this case, in particular regarding the position of the territorial principle, the relationship between the spatial and personal models of jurisdiction, the use of lethal force and the dividing and tailoring of Convention rights. This article therefore proposes that the existing models be replaced with a cause-and-effect model of jurisdiction triggered on the basis of (i) cause in fact; and (ii) an assessment of the scope of responsibility based on the principles of foreseeability and remoteness. This article further proposes

* Merton College, University of Oxford. I am grateful to Emilie McDonnell and to the OUULJ editorial team for their insightful comments made on earlier drafts. All errors remain my own.

that the Convention rights be divided and tailored such that the extent of States' obligations be proportionate to the capacity of States to fulfill these obligations, based on an assessment of the *de facto* situation.

Introduction

The scope of the application of the European Convention on Human Rights (ECHR or Convention) is governed by Article 1, which requires the High Contracting Parties to ‘secure to everyone within their jurisdiction the rights and freedoms defined in Section I of the Convention’.¹ While neither Article 1 nor any other part of the ECHR provide a definition of the term ‘jurisdiction’, the reference to the concept of ‘within their jurisdiction’, rather than ‘within their territory’, indicates that States may be obliged to secure rights and freedoms to those outside their territory, such that the ECHR has extraterritorial application. The importance of establishing the jurisdiction of a particular State lies in the fact that jurisdiction ‘is a threshold criterion. The exercise of jurisdiction is a necessary condition for a Contracting State to be able to be held responsible for acts or omissions imputable to it’.² Thus, if a potential claimant is not within the jurisdiction of a State, then the State cannot be held responsible, and the application is inadmissible *ratione personae*³ before the European Court of Human Rights (the Court). The seminal case of *Al-Skeini and others v United Kingdom*⁴ (*Al-Skeini*) is a restatement of the principles governing the scope of the extraterritorial application of the ECHR. As such, this article will first assess the implications of *Al-Skeini*, which explicitly

¹ Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights, as amended) (ECHR), Art 1.

² *Al-Skeini and Others v UK* [GC] (2011) 53 EHRR 18 [130].

³ Directorate of the Jurisconsult, ‘Practical Guide on Admissibility Criteria’ (Council of Europe/European Court of Human Rights, 30 April 2020)
<https://www.echr.coe.int/Documents/Admissibility_guide_ENG.pdf> accessed 10 May 2021.

⁴ *Al-Skeini* (n 2).

recognised the personal model of jurisdiction, accepted that Convention rights can be divided and tailored and abandoned the concept of *espace juridique* (Part 1). Then, this article will assess the uncertainties that have remained following *Al-Skeini* (Part 2). Finally, this article will argue that the cause-and-effect model of jurisdiction should be adopted to determine the scope of the extraterritorial application of the ECHR (Part 3), such that an individual who is causally affected by an act or omission of a State is within its jurisdiction. The adoption of the cause-and-effect model serves to ensure that States are not ‘gentlemen at home, hoodlums elsewhere’,⁵ that is, that States do not escape responsibility for violations committed on the territory of another State, which they could not perpetrate on their own territory.

1. Analysing Al-Skeini

Al-Skeini concerned the deaths of six Iraqi civilians who had been killed or fatally wounded by British soldiers in Basra (southern Iraq) in 2003, when the United Kingdom was an occupying power. Four of them were shot by British soldiers, another victim had been beaten by British soldiers and then forced to jump into a river where he drowned, and 93 wounds had been found on the body of the last victim who had died while detained at a British military base.

Their relatives contended that the victims were within the jurisdiction of the United Kingdom under Article 1 and that the United Kingdom had breached its procedural obligation under Article 2 in failing to carry out an adequate and effective

⁵ *ibid* [18] (Judge Bonello).

investigation into the deaths. The Court held that the victims fell within the jurisdiction of the United Kingdom under Article 1 and that there had been a breach of the procedural obligation under Article 2. The next part of this article will assess: (A) the treatment of the spatial model and personal model of jurisdiction; (B) Convention rights being able to be divided and tailored; and (C) the concept of *espace juridique*, in *Al-Skeini*.

A. Spatial Model & Personal Model of Jurisdiction

Prior to *Al-Skeini*, the extraterritorial application of the ECHR was premised on two strands of jurisprudence on jurisdiction: the spatial model and the personal model.⁶ The spatial model, developed in *Loizidou v Turkey*,⁷ dictates that jurisdiction arises when a State ‘exercises effective control of an area outside its national territory’⁸ in which the affected individual is situated. Meanwhile, the personal model lays down that jurisdiction arises whenever a State exercises authority and control over an individual.⁹

The Court in *Al-Skeini* explicitly recognised the personal model of jurisdiction, stating that ‘in certain circumstances, the use of force by a State’s agents operating outside its territory may

⁶ Marko Milanovic, ‘*Al-Skeini* and *Al-Jedda* in Strasbourg’ (2012) 23 (1) EJIL 121.

⁷ *Loizidou v Turkey* (1996) 21 EHRR 188.

⁸ *ibid* [62].

⁹ The Commission declared that ‘authorised agents of a State, including diplomatic or consular agents bring other persons or property within the jurisdiction of that State to the extent that they exercise authority over such persons or property. Insofar as, by the acts or omissions, they affect such persons or property, the responsibility of the State is engaged’: *Cyprus v Turkey* [1975] ECHR 3.

bring the individual thereby brought under the control of the State's authorities into the State's Article 1 jurisdiction'.¹⁰ The Court emphasised that 'what is decisive in such cases is the exercise of physical power and control over the person in question'.¹¹

The Court retained the requirement of an exercise of public powers by the State in order to trigger jurisdiction, stating that the extraterritorial jurisdiction of a State is established where 'through the consent, invitation or acquiescence of the Government of that territory, it exercises all or some of the public powers normally to be exercised by that Government'.¹² The Court thereby held that there was a jurisdictional link between the United Kingdom and the victims, as 'the United Kingdom ... assumed in Iraq the exercise of some of the public powers normally to be exercised by a sovereign government', in particular assuming 'authority and responsibility of the maintenance of security in south-east Iraq'.¹³ As Milanovic has noted, '*a contrario*, had the UK *not* exercised such public powers, the personal model of jurisdiction would not apply'.¹⁴ Thus, the Court applied an intermediate model somewhere between the spatial model and the personal model by emphasising that the State had exercised public powers as well as authority and control over the victims.

In explicitly recognising the personal model, *Al-Skeini* departed from the position of the Court in *Banković and Others v Belgium and Others*¹⁵ (*Banković*). *Banković* concerned the deaths and injuries resulting from a missile strike by NATO aircraft on the

¹⁰ *Al-Skeini* (n 2) [136].

¹¹ *ibid* [136].

¹² *ibid* [135].

¹³ *ibid* [149].

¹⁴ Milanovic (n 6).

¹⁵ *Banković and Others v Belgium and Others* (dec.) [GC] [2001] ECHR 890.

Serbian radio and television station in Belgrade in the then Federal Republic of Yugoslavia in 1999. Although the decision of the 17-member Grand Chamber of the Court was unanimous, it has attracted much academic criticism. The Court in *Banković* rejected the applicants' interpretation of its earlier jurisprudence as constituting recognition of a separate basis for extraterritorial jurisdiction amounting to the personal model¹⁶ and reaffirmed the spatial model as an exception to the primarily territorial application of the ECHR, stating that extraterritorial jurisdiction arises where a State 'through the effective control of the relevant territory and its inhabitants abroad ... exercises all or some of the public powers normally to be exercised by that Government'.¹⁷ Thus, in confirming that jurisdiction can arise where a State agent exercises authority and control over an individual, *Al-Skeini* departed from the spatial model adopted in *Banković*.

While *Al-Skeini* has been described as representing a 'potentially massive expansion of the scope of the application of the Convention'¹⁸ in light of its departure from the position in *Banković*, its explicit recognition of the personal model does not depart from the post-*Banković* line of authority in support of the personal model. Four cases which demonstrate the substantive shift to the personal model will now be outlined. In *Issa and Others v Turkey*¹⁹ (*Issa*), the Court rejected the Turkish government's submissions based on *Banković* and stated that 'a State may also be held accountable for violation of the Convention rights and freedoms of persons who are in the territory of another State but who are found to be under the former State's authority and control through its agents operating—whether lawfully or

¹⁶ *ibid* [81].

¹⁷ *ibid* [71].

¹⁸ *Al-Saadoon & Others v Secretary of State for Defence* [2016] EWCA Civ 811 [33].

¹⁹ *Issa and Others v Turkey* (2004) 41 EHRR 567.

unlawfully—in the latter State'.²⁰ Thus, had it been established (which on the facts it was not) that Turkish armed forces conducted operations in the area in question²¹ and had detained and killed the victims, the victims would have been within Turkey's jurisdiction by virtue of the personal model. In *Öcalan v Turkey*²² (*Öcalan*), the Court accepted that the applicant came within Turkey's jurisdiction on account of being under the authority and control of Turkish security forces following his arrest inside an aircraft in Nairobi Airport, noting that 'it is common ground that, directly after being handed over to the Turkish officials by the Kenyan officials, the applicant was effectively under Turkish authority and therefore within the "jurisdiction" of that State'.²³ In *Pad and Others v Turkey*²⁴ (*Pad*), the Court reiterated the personal model as espoused in *Issa*²⁵ and stated that, had the Turkish government not already admitted that the fire discharged from its helicopters caused the killing of the victims and had it been disputed by the parties that the victims came within the jurisdiction of Turkey, it would have applied the personal model as espoused in *Issa* to assess jurisdiction.²⁶ Finally, in *Isaak and Others v Turkey*²⁷ (*Isaak*) the Court declared admissible a complaint regarding the death of an individual at the hands of Turkish Cypriot police in the United Nations buffer zone in Cyprus. Despite taking place in an area over which Turkey did not have effective overall control, the Court found that the victim had been 'under the authority and/or effective control of the

²⁰ *ibid* [71].

²¹ *ibid* [81].

²² *Öcalan v Turkey* [GC] (2005) 41 EHRR 985.

²³ *ibid* [91].

²⁴ *Pad and Others v Turkey* App no 60167/00 (ECtHR, 28 June 2007).

²⁵ *ibid* [53].

²⁶ *ibid* [54].

²⁷ *Isaak and Others v Turkey* (dec.) App no 44587/98 (ECtHR, 24 September 2008).

respondent state through its agents'.²⁸ In *Al-Skeini* itself, the Court stated that its previous cases were not solely about control over particular geographical areas, rather concerned the exercise of physical power and control over the person in question.²⁹ Thus, it is evident that rather than presenting a significant expansion of the scope of the application of the ECHR in explicitly recognising the personal model, the Court in *Al-Skeini* sought to restore order following the development of two competing strands of jurisprudence on jurisdiction. Therefore, the *Banković* oversight—where the Court had failed to properly consider the personal model—was corrected.

Notwithstanding the body of preceding jurisprudence and *Al-Skeini*'s explicit recognition of the personal model, vestiges of the *Banković* approach remain in *Al-Skeini* in three respects. Firstly, the Court retained the requirement of an exercise of public powers by the State in order to trigger jurisdiction. 'Thus, where ... authorities of the Contracting State carry out executive or judicial functions on the territory of another State, the Contracting State may be responsible for breaches of the Convention thereby incurred'.³⁰ Jurisdiction does not arise merely on account of the exercise of force or any violation of an ECHR right, rather the exercise of public powers is required. Secondly, while the Court did recognise the personal model, it did not at any point explicitly overrule the spatial model espoused by *Banković*, such that it continues to be applicable. Thirdly, the Court reaffirmed the territorial principle, stating that 'a State's

²⁸ *ibid.*

²⁹ *Al-Skeini* (n 2) [136] 'The Court does not consider that jurisdiction in the above cases arose solely from the control exercised by the Contracting State over the buildings, aircraft or ship in which the individuals were held. What is decisive in such cases is the exercise of physical power and control over the person in question'.

³⁰ *Al-Skeini* (n 2) [135].

jurisdictional competence under Article 1 is primarily territorial' and that 'acts of the Contracting States performed, or producing effects, outside their territories can constitute an exercise of jurisdiction within the meaning of Article 1 only in exceptional cases',³¹ requiring special justification in the particular circumstances of each case. Thus, as Milanovic has noted, the Court retained 'the basic *Banković* posture that recognition of extraterritorial jurisdiction must remain exceptional'.³²

B. Dividing and Tailoring of Convention Rights

The Court in *Al-Skeini* overturned the Court's ruling in *Banković* in holding that Convention rights can be divided and tailored. In *Banković*, the applicants argued that 'the determination of 'jurisdiction' can be carried out by adapting the 'effective control' criteria ... so that the extent of the positive obligation under Article 1 of the Convention to secure Convention rights would be proportionate to the level of control in fact exercised'.³³ In response to this, the Court stated that 'the wording of Article 1 does not provide any support for the applicants' suggestion that the positive obligation in Article 1 ... can be divided and tailored in accordance with the particular circumstances of the extra-territorial act in question',³⁴ such that either all or none of the Convention rights applied. The Court in *Al-Skeini* stated that where a State exercises effective control of an area, 'the controlling State has the responsibility under Article 1 to secure, within the area under its control, the entire range of substantive rights set out in the Convention and those additional Protocols which it has ratified'.³⁵ Conversely, where a State through its

³¹ *ibid* [131].

³² Milanovic (n 6).

³³ *Banković* (n 15) [46].

³⁴ *ibid* [75].

³⁵ *Al-Skeini* (n 2) [138].

agents exercises authority and control over an individual, ‘the State is under an obligation under Article 1 to secure to that individual the rights and freedoms under Section I of the Convention that are relevant to the situation of that individual’.³⁶ Holding that Convention rights can be divided and tailored allowed the Court to recognise the personal model, which had been rejected in *Banković*. As Lord Hope has noted, ‘the concept of dividing and tailoring goes hand in hand with the principle that extraterritorial jurisdiction can exist whenever a state through its agents exercises authority and control over an individual. The court need not now concern itself with the question of whether the state is in a position to guarantee Convention rights to that individual other than those it is said to have breached’.³⁷

C. Espace Juridique

The Court also departed from *Banković* in abandoning the concept of *espace juridique*. As articulated in *Banković*, *espace juridique* dictates that the operation of the ECHR is limited to the territorial borders of its Contracting States and does not apply beyond this legal space. In *Banković*, the Court resisted the extraterritorial application of the ECHR by emphasising the fact that ‘the Convention is a multi-lateral treaty operating ... in an essentially regional context and notably in the legal space (*espace juridique*) of the Contracting States. The [Federal Republic of Yugoslavia] clearly does not fall within this legal space’.³⁸ Thus, the Court stressed that the ECHR was not designed to be applied throughout the world, even in respect of the conduct of Contracting States. However, in subsequent cases the Court did not apply the concept of *espace juridique*, with the ECHR being

³⁶ *ibid* [137].

³⁷ *Smith v Ministry of Defence* [2014] AC 52 [49].

³⁸ *Banković* (n 15) [80].

applicable to the alleged acts of Turkish agents in the non-Contracting States' territories of northern Iraq,³⁹ Kenya,⁴⁰ Iran,⁴¹ and the United Nations buffer zone in Cyprus.⁴² The ECHR has also been applied in relation to the acts of the Portuguese authorities in international waters.⁴³ In *Al-Skeini*, the Court overruled *Banković* holding that 'the importance of establishing the occupying State's jurisdiction' does not imply 'that jurisdiction under Article 1 of the Convention can never exist outside the territory covered by the Council of Europe Member States'.⁴⁴ Thus, *espace juridique* has finally been laid to rest, such that ECHR obligations can theoretically arise in any part of the world.

Thus, while the Court did not explicitly overrule *Banković* and retained the public powers requirement, there was an expansion in the scope of the extraterritorial application of the ECHR through explicitly recognising the personal model, accepting that Convention rights can be divided and tailored and abandoning the concept of *espace juridique*.

2. Uncertainties Following *Al-Skeini*

Despite the clarifications offered by *Al-Skeini*, some uncertainties have remained regarding (A) the position of the territorial principle; (B) the relationship between the spatial and personal

³⁹ *Issa* (n 19).

⁴⁰ *Öcalan* (n 22).

⁴¹ *Pad* (n 24).

⁴² *Isaak* (n 27).

⁴³ *Women on Waves and Others v Portugal* App no 31276/05 (ECtHR, 3 February 2009).

⁴⁴ *Al-Skeini* (n 2) [142].

models; (C) the use of lethal force; and (D) the dividing and tailoring of Convention rights.

A. Position of the Territorial Principle

The position of the territorial principle⁴⁵ is uncertain after the Court broadened the scope of the extraterritorial application of the ECHR. The Court's references to the territorial principle appear to be indicative of a strong presumption in favour of jurisdiction being territorial. However, according to Duttwiler, the territorial principle is 'to be taken with a grain of salt', as 'the reality is that the Convention organs have applied the ECHR to numerous extraterritorial situations which occurred on the high seas and on four different continents'.⁴⁶ Similarly, Lord Hope has argued that the word 'exceptional' is included 'not to set an especially high threshold for circumstances to cross before they can justify a finding that the state was exercising jurisdiction extraterritorially', but only 'to make it clear that, for this purpose, the normal presumption that applies throughout the state's territory does not apply'.⁴⁷ Lord Dyson has gone so far as to extrajudicially state that the 'fundamental principle is that of the exercise of control and authority', such that 'the territorial principle loses its special significance'.⁴⁸ According to His Lordship, 'it is clearer simply to say that, whenever the state

⁴⁵ *Banković* (n 15) [61].

⁴⁶ Michael Duttwiler, 'Authority, Control and Jurisdiction in the Extraterritorial Application of the European Convention on Human Rights' (2012) 30 (2) *Netherlands Quarterly of Human Rights* 137.

⁴⁷ *Smith* (n 37) [30].

⁴⁸ Lord John Dyson, 'The Extraterritorial Application of the European Convention on Human Rights: Now on a Firmer Footing, but is it a Sound One?' (Judiciary of England and Wales, 30 January 2014) <<https://www.judiciary.uk/wp-content/uploads/JCO/Documents/Speeches/lord-dyson-speech-extraterritorial-reach-echr-300114.pdf>> accessed 18 May 2021.

exercises control and authority over an individual, it is under an obligation under Article 1 to secure the rights and freedoms of the Convention to that individual wherever he or she happens to be.⁴⁹

B. Relationship Between the Spatial and Personal Models

Al-Skeini chronicles the fact that there are two principal bases for establishing extraterritorial jurisdiction—effective control over an area and State agent authority and control—without elucidating the relationship between the two. If a State exercises effective control over an area, it can be assumed that its agents also have authority and control over the people living in that area. The opposite is not true, as authority and control over particular individuals does not necessarily correlate with effective control over the whole area. This lack of clarity is exacerbated by the fact that, in *Al-Skeini*, the Court retained the public powers requirement from *Banković* as an element of the personal model, referring to paragraph 71 of *Banković* as support.⁵⁰ It is strange that the Court referenced this paragraph, which concerned the spatial rather than the personal model, stating that jurisdiction arises ‘when the respondent State, through the effective control of the relevant territory and its inhabitants abroad . . . exercises all

49 *ibid.* This notion is corroborated by Barbara Miltner, who has argued that ‘in the extraterritorial context, it is the state’s nexus to persons that becomes the pivotal element of Article 1 jurisdiction’ and that ‘jurisdictional connections *ratione personae* have become the necessary and sufficient means for establishing Article 1 jurisdiction’: Barbara Miltner, ‘Revisiting Extraterritoriality after *Al-Skeini*: The ECHR and Its Lessons’ (2012) 33 (4) *Michigan Journal of International Law* 693.

⁵⁰ *Al-Skeini* (n 2) [135].

or some of the public powers normally to be exercised by that Government'.⁵¹

Moreover, as Raible has argued, 'the geographical model collapses into the personal one when it is applied to smaller and smaller areas',⁵² such as mere places or vessels. For instance, *Medvedyev and others v France*⁵³ concerned France having full and exclusive control over a Cambodian merchant ship and its crew.⁵⁴ Meanwhile, *Al-Saadoon and Mufdhi v United Kingdom*⁵⁵ concerned United Kingdom authorities exercising control over the premises of a British-run prison in Iraq.⁵⁶ The lack of distinction between the two tests beyond merely a difference in scale has continued in later case law, with the Court in *Jaloud v the Netherlands*⁵⁷ (*Jaloud*) explicitly mentioning both models but making no distinction between them in application. Thus, the relationship and delineation between the spatial and personal model has remained unclear, creating uncertainty as to the prevailing test of extraterritorial jurisdiction.

C. Use of Lethal Force

The explicit recognition of the personal model of jurisdiction in *Al-Skeini* has raised the question of the extent to which this model captures the use of lethal force. Logically, killing alone constitutes an exercise of authority and control over an individual by totally

⁵¹ *Banković* (n 15) [71].

⁵² Lea Raible, 'The Extraterritoriality of the ECHR: Why *Jaloud* and *Pisari* Should Be Read as Game Changers' [2016] 2 European Human Rights Law Review 161.

⁵³ *Medvedyev and Others v France* [GC] (2010) ECHR 384.

⁵⁴ *ibid* [66]-[67].

⁵⁵ *Al-Saadoon and Mufdhi v United Kingdom* (2010) 51 EHRR 9.

⁵⁶ *ibid* [88].

⁵⁷ *Jaloud v The Netherlands* [GC] App no 47708/08 (ECtHR, 20 November 2014).

removing his autonomy. This notion has been corroborated by Judge Albuquerque, who stated that ‘the shooting of an individual by State agents constitutes the ultimate form of the exercise of State control’.⁵⁸

As Milanovic has reasoned, *Banković* and *Al-Skeini* indicate that ‘while the ability to kill is “authority and control” over the individual if the state has public powers, killing is not authority and control if the state is merely firing missiles from an aircraft’.⁵⁹ Meanwhile, *Issa, Pad* and *Isaak* indicate that killing, including by bombing from a helicopter, constitutes an exercise of authority and control over an individual triggering jurisdiction even in the absence of an exercise of public powers. In particular, as Milanovic has noted, the application of the personal model to a fact pattern similar to *Banković*—a killing taking place by missile fire from an aircraft—renders *Pad* directly contradictory to *Banković*.⁶⁰

The question of the extent to which the personal model captures the use of lethal force is further complicated by the case of *Georgia v Russia* (II),⁶¹ where the Court held that Russia did not have jurisdiction under either the spatial⁶² or the personal model

⁵⁸ *Georgia v Russia* (II) App no 38263/08 (ECtHR, 21 January 2021) [9] (Judge Albuquerque).

This has been further corroborated by Leggatt J who stated that killing would give rise to jurisdiction under the personal model as ‘it is the ultimate exercise of physical control’: *Al-Saadoon* (n 18) [95].

⁵⁹ Milanovic, ‘*Al-Skeini* and *Al-Jedda* in Strasbourg’ (n 6).

⁶⁰ *ibid.*

⁶¹ *Georgia v Russia* (n 58).

⁶² The Court held that Russia did not have jurisdiction under the spatial model, as armed confrontation and fighting between enemy military forces creates a situation of chaos where there is no control over an area. According to the Court, ‘in the event of military operations—including, for example, armed attacks, bombing or

for the purpose of assessing alleged substantive violations of Article 2 during the active phase of hostilities in an international armed conflict (IAC),⁶³ such that serious allegations of unlawful killings of civilians during the first five days of conflict were inadmissible.⁶⁴ In assessing the applicability of the personal model, the Court sought to distinguish this case from prior cases where lethal force was sufficient to create a jurisdictional link, such as *Issa*, stating that:

[T]hose cases concerned isolated and specific acts involving an element of proximity. By contrast, the active phase of hostilities which the Court is required to examine in the present case in the context of an international armed conflict is very different, as it concerns bombing and artillery shelling by Russian armed forces.⁶⁵

While the Court appears to be excluding heavier weaponry or methods causing larger scale damage from jurisdiction, it is difficult to enumerate the uses of force that are considered ‘isolated and specific acts involving an element of

shelling—carried out during an international armed conflict one cannot generally speak of ‘effective control’ over an area. The very reality of armed confrontation and fighting between enemy military forces seeking to establish control over an area in a context of chaos means that there is no control over an area’: *Georgia v Russia* (n 58) [126].

⁶³ *Georgia v Russia* (n 58) [137]-[138].

⁶⁴ In spite of this conclusion, the Court adopted a less restrictive approach than the one in *Banković* by holding that: (i) Russia had jurisdiction over individuals who were detained and subjected to ill-treatment even during the active phase of hostilities: *ibid* [239], [269]; (ii) Russia had jurisdiction for the purpose of investigating potential substantive violations of Article 2 even if the deaths occurred during the active phase of hostilities: *ibid* [331].

⁶⁵ *ibid* [131]-[132].

proximity'. As Milanovic has commented, it is unclear whether 'hand to hand, gun to gun, tank to tank, drone to target'⁶⁶ forms of force fulfill this requirement. Moreover, as the Court confined its holding to IACs only, it is unclear how lethal force in an extraterritorial non-international armed conflict or in circumstances below the armed conflict threshold would be treated. In light of these issues, the extent to which the use of lethal force is captured by the personal model remains unclear.

D. Dividing and Tailoring of Convention Rights

The implications of the conclusion in *Al-Skeini* that Convention rights can be divided and tailored remain to be worked out, as it is unclear *in what manner* and *at what time* they can be so divided and tailored. While the Court in *Al-Skeini* did indicate that Article 2 should be interpreted flexibly without imposing unrealistic burdens, on account of the conditions in Iraq being far more difficult than those in the United Kingdom in peacetime,⁶⁷ *Al-Skeini* and the following cases have offered little specific guidance as to how rights are to be divided and tailored. For instance, Leggatt J in *Al-Saadoon and Others v Secretary of State for Defence* described the fact that rights could be divided and tailored as the 'critical point of departure',⁶⁸ but failed to offer any particular guidance. This matter is particularly important as, while it may appear obvious that an individual's right to life should be protected, the stringency of the Court's review of State action to secure rights, as well as the potential inclusion of particular

⁶⁶ Marko Milanovic, 'Georgia v. Russia No. 2: The European Court's Resurrection of *Banković* in the Contexts of Chaos' (*EJIL: Talk!*, 25 January 2021) <<https://www.ejiltalk.org/georgia-v-russia-no-2-the-european-courts-resurrection-of-bankovic-in-the-contexts-of-chaos/>> accessed 18 May 2021.

⁶⁷ *Al-Skeini* (n 2) [168]-[177].

⁶⁸ *Al-Saadoon* (n 18) [57].

Articles, such as the Article 13 right to marry, that appear out of place in the context of extraterritorial jurisdiction, require nuanced and principled reasoning. The contentious nature of this matter was evident in *Jaloud*, where the Court was unanimous in finding jurisdiction and a violation of the procedural obligation under Article 2 but there was a 10-7 split amongst the judges regarding the flexibility that States should be accorded in extraordinary situations such as occupation and armed conflict and the stringency of the Court's review of State action in this context.

3. *The Cause-and-Effect Model of Jurisdiction*

The uncertainties following *Al-Skeini* indicate that the jurisprudence on extraterritorial jurisdiction remains 'bedevilled by an inability or an unwillingness to establish a coherent and axiomatic regime, grounded in essential basics and even-handedly applicable across the widest spectrum of jurisdictional controversies'.⁶⁹ In light of this, it is necessary to determine the proper scope of the extraterritorial application of the ECHR, having regard to the purpose of Article 1 and the ECHR as a whole.

This article proposes that the spatial and personal models be replaced with a cause-and-effect model of jurisdiction, such that an individual causally affected by an act or omission of a State will be within its jurisdiction. Jurisdiction would be triggered on

⁶⁹ *Al-Skeini* (n 2) [4] (Judge Bonello).

the basis of a two-tier test: (i) cause in fact, and (ii) an assessment of the scope of responsibility.⁷⁰ Cause in fact seeks to establish the causal relationship between the act or omission of the defendant State and the harmful outcome through the ‘but for’ test. The ‘but for’ test entails the question of whether the harmful outcome would have occurred but for the preceding act or omission of the defendant State. This test indicates the outer limits of liability, as it yields a potentially infinite number of causes for any given harm. An assessment of the scope of responsibility is required to restrict liability, as it would be unreasonable to hold a defendant State liable for harmful results beyond their control. As such, the second limb of the cause-and-effect model seeks to determine which consequences of their wrongful acts the defendant State should be responsible for.⁷¹

An assessment of the scope of responsibility involves the principles of foreseeability and remoteness. The principle of foreseeability requires that the wrongful harm could have been foreseen as a consequence of the act or omission by a reasonable person at the time it was carried out, while the principle of remoteness operates to exculpate States where there is insufficient proximity between the State’s act or omission and the harm. The cause-and-effect model does not amount to ‘anyone adversely affected by an act imputable to a Contracting State, wherever in the world that act may have been committed or its consequences felt’⁷² being brought within the jurisdiction of that State, as it includes the principles of foreseeability and remoteness that will

⁷⁰ Markus Kellner and Isabelle Durant, ‘Causation’ in Atilla Fenyves et al. (eds), *Tort Law in the Jurisprudence of the European Court of Human Rights* (De Gruyter 2012) 449.

⁷¹ Ilias Plakokefalos, ‘Causation in the Law of State Responsibility and the Problem of Overdetermination: In Search of Clarity’ (2015) 26(2) *The European Journal of International Law* 471.

⁷² *Banković* (n 15) [75].

exculpate a State in spite of the ‘but for’ test establishing a factual causal link.

In advocating in favour of a cause-and-effect model of jurisdiction, the next part of this article will (A) discuss the case law of the Court as it relates to the cause-and-effect model, (B) explore the manner in which the cause-and-effect model accords with the principle of the universality of human rights and the living instrument interpretive approach, (C) address arguments against the adoption of the cause-and-effect model, and (D) briefly outline the appropriate application of the notion that Convention rights can be divided and tailored.

A. Case Law

Despite the rejection of a cause-and-effect notion of jurisdiction by the Court in *Banković*,⁷³ there is precedent for its application. *Andreou v Turkey*⁷⁴ concerned the cross-border shooting of a demonstrator situated on Greek-Cypriot territory by Turkish armed forces situated on Turkish-Cypriot territory. The Turkish government argued that the affected individual did not fall under its jurisdiction, as it lacked control over the United Nations buffer zone or the Greek-Cypriot National Guard ceasefire line. The Court disagreed, stating that:

[E]ven though the applicant sustained her injuries in territory over which Turkey exercised no control, the opening of fire on the crowd from close range, which was the *direct and immediate cause* of those injuries, was such that

⁷³ *ibid* [75].

⁷⁴ *Andreou v Turkey* App no 45653/99 (ECtHR, 27 October 2009).

the applicant must be regarded as ‘within [the] jurisdiction’ of Turkey within the meaning of Article 1.⁷⁵

Thus, the Court focussed on the causal link between the actions of State agents and the harmful outcome.

In addition to this particular case, the cause-and-effect model is not inconsistent with the current trend of the case law adopted by the Court. As has been explored above, there is a line of authority—including *Issa*, *Öcalan*, *Pad* and *Isaak*—that supports the personal model without a public powers requirement. In *Hassan v United Kingdom*, the Court recognised that this case concerned a period ‘before the United Kingdom and its coalition partners had declared that the active hostilities phase of the conflict had ended and that they were in occupation, and before the United Kingdom had assumed responsibility for the maintenance of security in the South East of the country’.⁷⁶ Thus, regardless of the fact that United Kingdom had not yet exercised the very public powers referred to in *Al-Skeini*, the Court concluded that the United Kingdom had jurisdiction over the victim as he had been ‘within the physical power and control of the United Kingdom soldiers’⁷⁷ when he had been detained.

The case of *Jaloud* provides further support for the notion that the Court is moving toward adopting the cause-and-effect model. As Raible has commented, in *Jaloud*, the Court oscillated between the application of the personal and spatial model⁷⁸ and did not specify which model it ultimately applied. The Court sought to determine the relationship between Dutch troops and

⁷⁵ *ibid* [25] (emphasis added).

⁷⁶ *Hassan v UK* [GC] App no 29750/09 (ECtHR, 16 September 2014) [75].

⁷⁷ *ibid* [76].

⁷⁸ Raible (n 52).

the UK armed forces,⁷⁹ which would seem to indicate that the Court was trying to establish that the Netherlands had control over the area despite not being the occupying power at the time, and thereby establish jurisdiction. However, the Court held that the death of Jaloud occurred within the jurisdiction of the Netherlands, as it had asserted ‘authority and control over persons passing through the checkpoint’,⁸⁰ which is reminiscent of the personal model. As Haijer and Ryngaert have noted, in clarifying that individuals killed by shots from a checkpoint fall within the jurisdiction of the State controlling the checkpoint, the Court has opened the door for patrol jurisdiction: ‘jurisdiction over individuals injured by patrol brigades (after all, both checkpoints and patrols are not permanent establishments),’⁸¹ given that a foot patrol can be considered a moving checkpoint. Such a broad interpretation of State agent authority and control comes close to the cause-and-effect model, indicating that the Court is moving toward adopting this model without explicitly stating that it is doing so.

B. The Principle of the Universality of Human Rights & the Living Instrument Interpretive Approach

The cause-and-effect model of jurisdiction accords with the principle of the universality of human rights: the notion that individuals are endowed with equal human rights simply by virtue of being human, regardless of where they live and who they are. The agenda heralded in the Preamble of the EHCR is the

⁷⁹ *Jaloud* (n 57) [141]-[149].

⁸⁰ *ibid* [152].

⁸¹ Friederycke Haijer and Cedric Ryngaert, ‘Reflections on *Jaloud v. the Netherlands*’ [2015] 19 *Journal of International Peacekeeping* 174.

‘universal and effective recognition and observance’⁸² of fundamental human rights. Preventing States from escaping responsibility for human rights violations abroad would be in line with this aim of securing universal recognition of human rights. As Judge Bonello argued in *Al-Skeini*, “‘universal’ hardly suggests an observance parcelled off by territory on the checkerboard of geography’.⁸³ This notion is further echoed in *Issa* where the Court stated that ‘Article 1 of the Convention cannot be interpreted so as to allow a State party to perpetrate violations of the Convention on the territory of another State, which it could not perpetrate on its own territory’.⁸⁴ If human rights are regarded as inherently ‘human’, it would be doctrinally inappropriate to treat individuals as deserving of a lesser standard of human rights depending on their location and, thereby, allow States to escape responsibility for their actions. The lower jurisdictional standard of the cause-and-effect model satisfies this objective.

The cause-and-effect model also accords with the living instrument interpretive approach to the ECHR. This approach characterises the ECHR as a ‘living instrument which ... must be interpreted in light of present-day conditions’.⁸⁵ In response to this demand, the ECHR must be interpreted in light of the development in the use of technology by States, particularly the proliferation of surveillance, air strikes by unmanned aerial vehicles (drone strikes), and cyberattacks. States have increasingly engaged in mass surveillance regimes, with the United Kingdom’s surveillance regime involving the collection of data belonging to those situated outside its territory, thereby impacting their Article

⁸² Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights, as amended) (ECHR), Preamble.

⁸³ *Al-Skeini* (n 2) [9] (Judge Bonello).

⁸⁴ *Issa* (n 19) [71].

⁸⁵ *Tyrer v United Kingdom* (1978) Series A no 26 [31].

8 right to privacy. Drone strikes, which seriously endanger civilian lives contrary to Article 2, have been increasingly utilised as a method of warfare. British statistics indicate that in the four years of war against ISIS in Iraq and Syria from 2014-2018, Reaper drones were deployed on more than 2,400 missions: almost two a day on average.⁸⁶ States are increasingly incorporating active cyber operations into their military arsenals, with cyberattacks having the capacity to cause widespread ramifications akin to a traditional armed attack. Surveillance, drone strikes, and cyberattacks are instances of conduct engaged in by States that are controlled centrally within a State's territory but produce effects outside its territory. Such situations where a State agent has insufficient proximity of relationship to an individual to directly exert their authority and control but is nonetheless able to significantly impact upon their rights are excluded by the spatial and personal models. If the use of lethal force does not suffice to create a jurisdictional link extraterritorially on account of being an exercise of authority and control, then exfiltrating data, reading emails or hacking phones is unlikely to suffice either. In fact, the United Kingdom Investigatory Powers Tribunal has relied upon this *Banković* analogy to rule that individuals located outside the United Kingdom but subject to its surveillance are not within the jurisdiction of United Kingdom.⁸⁷ Thus, the means by which a Contracting State can harm the rights of the citizens of another state without having to physically enter the territory of the latter state have changed since the ECHR first conceived of jurisdiction. This is exactly the kind of development that the living

⁸⁶ Dan Sabbagh, 'Killer drones: how many are there and who do they target?' *The Guardian* (London, 18 November 2019) <<https://www.theguardian.com/news/2019/nov/18/killer-drones-how-many-uav-predator-reaper>> accessed 18 May 2021.

⁸⁷ *Human Rights Watch Inc & Others v The Secretary of State for the Foreign & Commonwealth Office & Others* [2016] 5 WLUK 352.

instrument interpretive approach ought to capture. A shift to the cause-and-effect model would ensure that Contracting States are unable to escape responsibility in such scenarios.

Similarly, the cause-and-effect model would ensure that States are held responsible for their complicity in human rights abuses abroad, particularly torture. As Jackson has noted, States have been involved in foreign human rights abuses by sharing intelligence, selling equipment used in interrogations, or providing technical support. In such cases, ‘the foreign state is wholly in control of its own territory; there is no question of the complicit state exercising effective control of an area outside of its own territory’.⁸⁸ Moreover, the authority and control rest with the foreign state, which is responsible, for instance, for an arrest and torture, such that the complicit State’s agents facilitate but do not themselves exercise authority and control over the victim. Thus, under the existing interpretation of jurisdiction, the victim of the foreign state’s acts does not fall within the complicit state’s jurisdiction, such that States may facilitate acts of torture or other human rights violations committed abroad by foreign agents without incurring responsibility. This development in state practice should be captured by the living instrument interpretive approach, with the adoption of the cause-and-effect model ensuring that there is no distinction between torture or other human rights violations committed on one’s own territory and those facilitated on the territory of another state.

⁸⁸ Miles Jackson, ‘Freeing Soering: The ECHR, State Complicity in Torture and Jurisdiction’ [2016] 27(3) *The European Journal of International Law* 817.

C. Addressing Arguments Against the Cause-and-Effect Model of Jurisdiction

The Court has interpreted extraterritorial jurisdiction under the ECHR by reference to the narrow, primarily territorial understanding of jurisdiction in international law, with the corollary that extraterritorial jurisdiction has been considered exceptional and requiring special justification, rather than being the natural consequence of States' actions abroad. The Court in *Banković* declared itself satisfied that 'from the standpoint of public international law, the jurisdictional competence of a State is primarily territorial. While international law does not exclude a State's exercise of jurisdiction extra-territorially, the suggested bases of such jurisdiction ... are, as a general rule, defined and limited by the sovereign territorial rights of the other relevant States'.⁸⁹ Thus, the Court held that Article 1 'must be considered to reflect this ordinary and essentially territorial notion of jurisdiction, other bases of jurisdiction being exceptional and requiring special justification'.⁹⁰ This approach has continued in the later case law.⁹¹ However, it has been questioned in legal commentary whether this reliance on the understanding of jurisdiction in international law is appropriate.⁹² The international law notion of jurisdiction refers to the 'competence of a State to make, adjudicate and enforce the law with regard to a certain

⁸⁹ *Banković* (n 15) [59].

⁹⁰ *ibid* [61].

⁹¹ See for example *Issa* (n 19) [67]; *Al-Skeini* (n 2) [131].

⁹² Alexander Orakhelashvili 'Restrictive Interpretation of Human Rights Treaties in the Recent Jurisprudence of the European Court of Human Rights' (2003) 14 EJIL 529; Marko Milanovic 'From Compromise to Principle: Clarifying the Concept of State Jurisdiction in Human Rights Treaties' (2008) 8 HRLR 411; Alexandra R uth and Mirja Trisch '*Banković v Belgium*' (2003) 97 AJIL 168.

situation—in short: to regulate conduct⁹³ without infringing upon other States' jurisdiction. As such, jurisdiction under international law allocates competences among States thereby seeking to reconcile their competing sovereignties. Jurisdiction under Article 1 of the ECHR serves a different function: it is a threshold criterion that delimits the group of individuals whose Convention rights have to be secured and thus, the scope of a State's duties to secure those rights. As such, extraterritorial jurisdiction 'is derived from factual control and does not depend on any legal authority to act'.⁹⁴ As Scheinin has argued, conflating both concepts of jurisdiction confuses the issue of the legal consequences of an exercise of authority and control abroad with the permissibility of a state exercising jurisdiction beyond its own territory.⁹⁵ The legitimacy of a State's actions under international law cannot be a condition for establishing jurisdiction under the ECHR. If legitimacy were a condition, then the ECHR would apply where States act within the confines of their competence under international law. As De Schutter has reasoned, such an interpretation would lead to 'the paradoxical result that a State acting beyond its ... powers as recognized under international law, could not be held responsible for the consequences resulting

⁹³ Duttwiler (n 46).

⁹⁴ Stuart Wallace, *The Application of the European Convention on Human Rights to Military Operations* (CUP 2019). This has been corroborated by Judge Loucaides who has stated that jurisdiction for the purpose of Article 1 means 'actual authority, that is to say, the possibility of imposing the will of the State on any person': *Ilaşcu and Others v Moldova and Russia* [GC] (2005) 40 EHRR 1030 (Judge Loucaides).

⁹⁵ Martin Scheinin, 'Extraterritorial Effect of the International Covenant on Civil and Political Rights' in Coomans and Kamminga (eds), *Extraterritorial Application of Human Rights Treaties* (Intersentia 2004).

from these acts under the treaties it has agreed to'.⁹⁶ The application of human rights law 'cannot depend on whether a State abuses its power within or outside the jurisdictional confines imposed by international law. To be effective [in protecting the individual against the abuse of State power], human rights have to apply wherever a State wields its power'.⁹⁷ Given that extraterritorial jurisdiction should not be interpreted by reference to the understanding of jurisdiction in international law, the territorial principle should be disregarded and a conception of extraterritorial jurisdiction that regards it as the natural consequence of States' actions abroad adopted.

It could be argued that the living instrument interpretive approach is not to be applied to Article 1, such that the aforementioned present-day conditions are irrelevant. In *Banković*, the Court appeared to imply that Article 1—unlike the ECHR's substantive provisions—was not to be interpreted as a living instrument in accordance with present-day conditions,⁹⁸ as 'the scope of Article 1 ... is determinative of the very scope of the Contracting Parties' positive obligations and, as such, of the scope and reach of the entire Convention system of human rights protection'.⁹⁹ In this way, the Court sought to limit the scope for developing a more expansive interpretation of jurisdiction in the future. However, in endorsing a more expansive interpretation of jurisdiction than that found by the Court in *Banković*, the Court has effectively been treating Article 1 as subject to the living instrument interpretive approach. As Lord Hope has noted, the words 'to date' in Paragraph 131 of *Al-Skeini* indicate that 'the list

⁹⁶ Olivier De Schutter 'Globalization and Jurisdiction: Lessons from the European Convention on Human Rights' (2006) 6 *Baltic Yearbook of International Law* 185.

⁹⁷ Duttwiler (n 46).

⁹⁸ *Banković* (n 15) [64]-[65].

⁹⁹ *ibid* [65].

of circumstances which may require and justify a finding that the state was exercising jurisdiction extraterritorially is not closed¹⁰⁰. Thus, the aforementioned present-day conditions that inform the living instrument interpretive approach are relevant.

It could also be argued that expanding the scope of the application of the ECHR through the cause-and-effect model may be inappropriate in light of the practical limitations of the Court. As the Court already suffers from considerable procedural delays, a low jurisdictional standard could lead to the Court being overburdened. Judge Nußberger has argued that ‘justice delayed is justice denied. If all are included, if the limits of the system are stretched in time, space, and substance, the system might be overloaded and stop functioning properly’.¹⁰¹ As Milanovic has noted, the decision in *Banković* was driven partly by the desire not to ‘open the floodgates of litigation by considering every individual against whom force was used as falling under the protection of the Convention’.¹⁰² Likewise, the Court in *Georgia v Russia (II)*¹⁰³ concluded that, having regard to ‘the large number of alleged victims and contested incidents, the magnitude of the evidence produced, the difficulty in establishing the relevant circumstances’, ‘it is not in a position to in a position to develop its case-law beyond the understanding of the notion of ‘jurisdiction’ as established to date’.¹⁰⁴ In light of these concerns,

¹⁰⁰ *Smith* (n 37) [30]. This notion has been corroborated by Miles Jackson, who argues that the living instrument interpretive approach points ‘towards a more expansive understanding of jurisdiction in Article 1’: Jackson (n 88).

¹⁰¹ Angelika Nußberger, ‘The Concept of ‘Jurisdiction’ in the Jurisprudence of the European Court of Human Rights’ (2012) 65 CLP 241.

¹⁰² Milanovic, ‘*Al-Skeini* and *Al-Jedda* in Strasbourg’ (n 6).

¹⁰³ *Georgia v Russia (II)* (n 58).

¹⁰⁴ *ibid* [141].

Judge Nußberger suggests that a restrictive interpretation of jurisdiction may be necessary to ensure the effectiveness of the Court. However, it is submitted that such practical concerns are insufficient to outweigh the significance of providing redress for human rights violations. Moreover, the application of a threshold that discounts rights violations abroad undermines the integrity of international human rights law. As Judge Nußberger has commented, accepting ‘that grave human rights violations remain unanswered’¹⁰⁵ might undermine the credibility of the Court. While selectivity is necessary to mitigate the increasing caseload of the Court and ensure its continuing effectiveness, this must be done on a reasoned basis through judgment on exemplary cases in pilot procedures, prioritisation of especially important cases, and a robust admissibility stage of proceedings.

Adjudicating on wartime measures adopted by States or on States’ military strategies might be regarded as being outside the Court’s institutional competence, such that jurisdiction should not arise in such circumstances. However, in the context of a conflict, the European Court of Human Rights may be the only available international institution to analyse human rights violations with binding effects for the parties. This is primarily on account of the fact that there is no permanent international court considering individual complaints regarding violations of international humanitarian law (IHL).¹⁰⁶ Even if there were a forum adjudicating on IHL, the protection granted under the ECHR extends to some rights and freedoms which are not

¹⁰⁵ Nußberger (n 101).

¹⁰⁶ Anastasiia Moiseieva, ‘The ECtHR in Georgia v. Russia—a farewell to arms? The effects of the Court’s judgment on the conflict in eastern Ukraine’ (*EJIL: Talk!*, 24 February 2021) <<https://www.ejiltalk.org/the-ecthr-in-georgia-v-russia-a-farewell-to-arms-the-effects-of-the-courts-judgment-on-the-conflict-in-eastern-ukraine/>> accessed 10 May 2021.

guaranteed or are not so widely covered under IHL, such as the freedom of the press, the right to assembly, strike or vote, as well as the right to private and family life.¹⁰⁷ Thus, the scenarios of invasions, occupations or other conflict are the exact sphere where the Court can demonstrate its unique competence.

D. Dividing and Tailoring of Convention Rights

The low jurisdictional standard of the cause-and-effect model of jurisdiction relies on Convention rights being able to be divided and tailored in order to assuage concerns of whether a State is actually in a position to guarantee all of the Convention rights to an individual. As such, the extent of States' obligations should be proportionate to the capacity of States to fulfill these obligations, based on an assessment of the *de facto* situation. Thus, jurisdiction should arise in regard to a certain Article when the performance of its obligations is within a State's control. This notion is echoed in the functional test of jurisdiction proposed by Judge Bonello in *Al-Skeini*. Judge Bonello argued that 'a Contracting State is obliged to ensure the observance of all those human rights which it is in a position to ensure'.¹⁰⁸ Similarly, according to Judges Yudkivska, Wojtyczek, and Chanturia, Article 1 should be read as stating 'a High Contracting Party shall secure the rights and freedoms defined in Section I of this Convention to everyone under its State power and the scope of the rights and freedoms to be secured should be adequate to the extent of the scope of effective State power'.¹⁰⁹

Necessitating that all territorial rights and obligations be available in the case of extraterritorial jurisdiction may be an

¹⁰⁷ *ibid.*

¹⁰⁸ *Al-Skeini* (n 2) [32] (Judge Bonello).

¹⁰⁹ *Georgia v Russia* (II) (n 58) [3] (Judges Yudkivska, Wojtyczek, and Chanturia).

impossible bar, especially given that particular Convention rights, such as the right to a fair trial, are predicated on the State having established the necessary institutions, which is difficult for an attacking or occupying force. However, it is necessary to prevent States from escaping all responsibility purely on account of being just an attacking or occupying force. For instance, there may be particular cases where an occupying force is so established in a particular country that it is reasonably able to provide greater protection than merely not engaging in killing or torture. As Judge Bonello has noted, ‘it is quite possible to envisage situations in which a Contracting State, in its role as an Occupying Power, has well within its authority the power not to commit torture or extrajudicial killings, to punish those who commit them and to compensate the victims—but at the same time that Contracting State does not have the extent of authority and control required to ensure to all persons the right to education or the right to free and fair elections’.¹¹⁰ Thus, the rights that a State is able to uphold would fall within its jurisdiction, while those it cannot uphold would not, based on a detailed assessment of the *de facto* situation.

Conclusion

While the Court did not explicitly overrule *Banković* and retained the public powers requirement, *Al-Skeini* has expanded the scope of extraterritorial application of the ECHR by explicitly recognising the personal model of jurisdiction, accepting that Convention rights can be divided and tailored, and abandoning the concept of *espace juridique*. Despite the clarification of the jurisprudence offered by this case, some uncertainties have

¹¹⁰ *Al-Skeini* (n 2) [32] (Judge Bonello).

remained, indicating the need for further clarification of the scope of the extraterritorial application of the ECHR. This article has advocated that such clarification could be achieved through the adoption of a cause-and-effect model of jurisdiction delimited by the principles of foreseeability and remoteness. In doing so, States are prevented from being ‘gentlemen at home, hoodlums elsewhere’ by being held responsible for their human rights violations abroad.

R (Dolan) v Secretary of State for Health and Social Care: Legality in the Time of Coronavirus

John Yap & Nicholas U Jin*

Abstract—The case of *R (Dolan) v Secretary of State for Health and Social Care* [2020] EWCA Civ 1605 concerned *inter alia* the legality of regulations imposed during England’s first national Covid-19 lockdown. Under challenge were the Health Protection (Coronavirus, Restrictions) (England) Regulations 2020, which were enacted by the Secretary of State for Health and Social Care on the 26th March 2020 under the Public Health (Control of Disease) Act 1984. Ultimately, the claim that the regulations were *ultra vires* failed in the Court of Appeal and permission to appeal to the Supreme Court was refused. This case comment critiques the Court of Appeal’s reasoning in dismissing the claim, and in particular, its treatment of the principle of legality. With respect, it argues that the Court’s strained and obfuscatory reasoning was analytically unsound and demonstrated undue deference to the executive. Taking the discussion of *Dolan* as the point of departure, this case comment will engage in the broader debate

* Mansfield College and Merton College, Oxford, respectively. We are immensely grateful to Professors Timothy Endicott and Rebecca Williams, as well as the dedicated team of editors at the OUULJ, for their comments on earlier drafts. All errors remain our own.

on legality, constitutionalism, and the role of the Courts in times of emergency. It concludes that principles of democratic legitimacy require the Courts to aspire to take a non-deferential business-as-usual approach to statutory construction and the principle of legality, even in a time of emergency.

Introduction

In times of emergency, the laws may have changed, but they speak the same language. At least, so said Lord Atkin in a dissenting judgement in *Liversidge v Anderson*,¹ which now stands as a proud monument in our legal history. Yet, when the case of *R (Dolan) v Secretary of State for Health and Social Care (Dolan)*²—challenging the legality of England’s first national lockdown—reached the Court of Appeal, the law appears to have spoken in a different tongue.

A. Background and Context

The Health Protection (Coronavirus, Restrictions) (England) Regulations 2020 (the Regulations)—imposing England’s first national lockdown in response to Covid-19—were enacted by the Secretary of State for Health and Social Care on the 26th March under the Public Health (Control of Disease) Act 1984 (the 1984 Act).

During the ‘emergency period’, Regulations 4 and 5 required certain types of business premises, as well as places of worship, to close. Public gatherings of more than two people were prohibited by Regulation 7. Most draconian was Regulation 6, which provided that: ‘no person may leave the place where they are living without reasonable excuse’. Regulation 8 provided for measures to enforce the lockdown and Regulation 9 made it a criminal offence to breach Regulations 4 through 8. Significantly, the Regulations applied to the population in England generally—subject only to narrowly-defined exceptions.

On the 29th and 30th October, *Dolan* came before the Court of Appeal (comprising Lord Burnett, and King and Singh

¹ [1942] AC 206 (HL) 244-247.

² [2020] EWCA Civ 1605, [2021] 1 All ER 780.

LJJ) seeking to challenge the legality of the Regulations. The applicants sought permission for judicial review on the basis that, *inter alia*, the Regulations were *ultra vires* the 1984 Act. The primary argument the applicants wished to advance was that the 1984 Act conferred power on the Secretary of State *only* to enact regulations targeting specific groups of individuals and premises—as opposed to the sweeping Regulations that were in fact enacted, which affected the entire population in general (the ‘scaling up’ argument). In addition, the applicants sought to argue that the Regulations fell afoul of the principle of legality; that they ought instead to have been made under Part 2 of the Civil Contingencies Act 2004 (the 2004 Act); and that Regulations 6 and 8 were *ultra vires* for their own particular reasons.

Permitting the applicants to bring a claim for judicial review on the *vires* of the Regulations,³ the Court of Appeal proceeded to dismiss it on its merits. It found instead that the Secretary of State did have power to enact the Regulations under the 1984 Act.⁴ On the 9th December, permission to appeal to the Supreme Court was refused on the basis that no arguable question of law was raised.

B. Approach and Method

This case comment will focus on the Court of Appeal’s reasoning concerning the *vires* of the Regulations. In particular, it will examine the Court’s treatment of the principle of legality in construing the Secretary of State’s power under the 1984 Act. With respect, we argue that the Court’s reasoning in *Dolan* was

³ *ibid* [42].

⁴ *ibid* [78]. The Court of Appeal at [115] upheld the decision of the High Court [2020] EWHC 1786, [2020] 7 WLUK 49 refusing judicial review on the applicants’ other public law and human rights grounds.

deeply problematic and serves as a negative exemplar of undue judicial deference in times of emergency.

The structure of this case comment is as follows. We begin in Part 1 by examining the legal issues at the heart of the applicants' challenge to the *vires* of the Regulations. Subsequently, in Parts 2 and 3, we lay out and critically analyse the judgement of the Court of Appeal, demonstrating that it strained unduly and erroneously to reach an executive-minded outcome. Finally, in Part 4, we engage more broadly in an evaluation of the various constitutional models for responding to emergencies and the role of the Courts therein.

We argue that, on the basis of authority and normative justification, Courts *must* resolve questions on the legality of executive action non-deferentially, even in times of emergency. It is *especially* in a climate of emergency—with its tendency to skew the exercise of statutory construction in favour of the executive—that Courts must aspire to take a business-as-usual approach to the principle of legality. This self-conscious posture would serve as a necessary corrective to protect fundamental rights, accord with principles of democratic legitimacy, and facilitate institutional *cooperation* within a model of 'legislative accommodation'—i.e. the dominant constitutional model in the United Kingdom in which emergency response is governed by law through an *ex ante* statutory framework.

1. Challenging the *Vires* of the Regulations

A. The Scaling Up Argument

The main way the applicants sought to challenge the *vires* of the Regulations was by the scaling up argument, which was essentially that the Secretary of State had power under the 1984 Act *only* to enact regulations targeting *specific* individuals and premises. Accordingly, they argued that the Regulations—which applied generally to the entire population of England—fell outside the four corners of that power.

The starting point of the rather intricate analysis was the Preamble of the Regulations, in which the Regulations were stated to have been enacted ‘in exercise of the powers conferred by Sections 45C(1), 3(c), 4(d), 45F(2) and 45P’ of the 1984 Act.

Section 45C(1) gave the Secretary of State the general power to ‘by regulations make provision for the purpose of preventing, protecting against, controlling or providing a public health response to the incidence or spread of infection’. The applicants submitted that the breadth of the general power in Section 45C(1) is *cut down* by Section 45C(3), which states that the ‘Regulations ... may in particular include’ three kinds of provisions. The first kind concerns requiring medical practitioners to record and notify suspected cases of infection; the second concerns conferring functions on local authorities to monitor public health risks. Only the third kind of provisions—those specified in Section 45C(3)(c)—was relevant, namely ‘imposing or enabling the imposition of restrictions or requirements on or in relation to persons, things or premises in the event of, or in response to, a threat to public health’.

This in turn, the applicants submitted, is *further cut down* by Section 45C(4), which specifies that that kind of provisions ‘include in particular’ four types of restrictions or requirements. Of the four specified, only Section 45C(4)(d) was stated as an enabling provision in the Preamble of the Regulations.

In other words, the applicants submitted that the Regulations were *necessarily* enacted by the Secretary of State under the power conferred by Section 45C(4)(d)—to impose ‘a special restriction or requirement’. A ‘special restriction or requirement’ is defined as ‘a restriction or requirement which can be imposed by a justice of the peace’.⁵

At this point, some context might be helpful. The power conferred on the Secretary of State to enact regulations under Section 45C was only added in 2008 by Section 129 of the Health and Social Care Act 2008 (the 2008 Amendments). Prior to that, a ‘special restriction or requirement’ could only be imposed by a justice of the peace if certain conditions were met.⁶ For orders made in relation to persons, a justice of the peace had to be satisfied that (a) the person may be infected, (b) the infection could present significant harm to human health, (c) there was a risk that the person might infect others, and (d) it was necessary to make the order to remove or reduce that risk.⁷ For orders made in relation to premises, a justice of the peace had to be satisfied that (i) the premises may be infected, (ii) the infection could present a significant harm to human health, (iii) there was a risk

⁵ Public Health (Control of Disease) Act 1984 (PH(CD)A 1984) s 45C(6)(a).

⁶ By virtue of PH(CD)A 1984 ss 45G(2), 45H(2) or 45I(2).

⁷ PH(CD)A 1984 s 45G(1).

that the premises might infect humans, and (iv) it was necessary to make the order to remove or reduce that risk.⁸

The applicants submitted that the Secretary of State could only impose a ‘special restriction or requirement’ in circumstances in which a similar order could have been made by a justice of the peace. Since it was clear that a justice of the peace could *only* impose ‘a special restriction or requirement’ on a case-by-case basis in relation to specific individuals or premises, the applicants submitted that enacting the Regulations fell outside the Secretary of State’s power under the 1984 Act.

Buttressing this argument, the applicants observed that the language of Section 45F— which supplements provisions concerning the Secretary of State’s power—seem to contemplate decisions being made on a specific and targeted basis. Section 45F(6) requires regulations enacted by the Secretary of State to ‘provide for a right of appeal to a magistrates’ court against *any decision* taken under the regulations by virtue of *a special restriction or requirement ... in relation to a person, thing or premises*’ (emphasis added). Likewise, Section 45F(7) makes reference to ‘any person, thing or premises’ and Section 45F(8) refers to the review of ‘*a special restriction or requirement*’ (emphasis added) in the singular grammatical form.

B. Supplementary Arguments

In addition to the scaling up argument, the applicants made a number of supplementary arguments.

⁸ *ibid* s 45I(1).

Firstly, the applicants ran an independent line of argument challenging the *vires* of Regulation 6.⁹ Briefly, their argument proceeded as follows.¹⁰ First, Regulation 6—which prohibited any person from leaving the place where one was living without a reasonable excuse—would have, without legal authority, amounted to the tort of false imprisonment.¹¹ Second, legislation may only authorise conduct amounting to a tort by express words or necessary implication.¹² Third, the 1984 Act did not, by express words or necessary implication, confer on the Secretary of State the power to impose home confinement on the general population. Therefore, Regulation 6 was *ultra vires*. Alternatively, the requirement to remain at home amounted to the imposition of ‘quarantine’, which fell outside the scope of the Secretary of State’s power under the 1984 Act.¹³

⁹ Philip Havers and Francis Hoar, ‘Statement of Facts and Grounds and Written Submissions of the Claimant’ (Published Court Document, 20 May 2020) [47].

¹⁰ This argument was originally advanced by Tom Hickman, Emma Dixon, and Rachel Jones in ‘Coronavirus and Civil Liberties in the UK’ (Blackstone Chambers, 6 April 2020) <<https://coronavirus.blackstonechambers.com/coronavirus-and-civil-liberties-uk/>> accessed 10 April 2021 [32]-[40]. See also, for a similar argument advanced by Lord Sandhurst and Benet Brandreth, ‘Pardonable in the Heat of Crisis—Building a Solid Foundation for Action’ (Society of Conservative Lawyers, 16 April 2020). <<https://www.conservativelawyers.com/publications>> accessed 10 April 2021, 4.

¹¹ *R (Jalloh) v Secretary of State for the Home Department* [2020] UKSC 4, [2021] AC 262, which held that a requirement backed by criminal sanctions that a person remain at home amounts to the common law tort of false imprisonment—even in the absence of physical restraint.

¹² *R (Gedi) v Secretary of State for the Home Department* [2016] EWCA Civ 409, [2016] 4 WLR 93.

¹³ PH(CD)A 1984 ss 45D(3) and 45G(2)(d).

Secondly, and rather similarly, the applicants ran an independent line of argument challenging the *vires* of Regulation 8.¹⁴ Regulation 8—which permits the use of reasonable force to remove a person in breach of Regulation 6 back to their place of residence—would, if exercised without legal authority, amount to the tort of trespass against the person. No provision within the 1984 Act, either expressly or by necessary implication, authorises the use of such force for enforcement. Therefore, ‘the enforcement powers [in Regulation 8] that allow for the use of force are *ultra vires*, irrespective of whether other parts of the Regulations are *vires*’.¹⁵

Thirdly, the applicants submitted that instead of the 1984 Act, the incredibly wide-ranging Regulations could *only* have been enacted under the 2004 Act:

The wide powers provided for under [the 2004 Act] are subject to strict limitations of time and rigorous Parliamentary scrutiny. Parliament, in passing [the 2008 Amendments], may be imputed to have in mind that any delegation of the power to make secondary legislation through the 1984 Act would supplement the delegated powers of [the 2004 Act]; and that powers that had the breadth of those delegated under [the 2004 Act] should only be used under that Act.¹⁶

¹⁴ This argument was originally advanced by Hickman, Dixon and Jones (n 10) [32]-[40], and Sandhurst and Brandreth (n 10) 3-6. Cited in Havers and Hoar (n 9) [47].

¹⁵ Havers and Hoar (n 9) [47].

¹⁶ *ibid.* For elaboration on the stringent safeguards found in the 2004 Act, see Keith Ewing, 'Covid-19: Government by Decree' [2020] 31(1) KCLJ 1, 14-15.

C. The Principle of Legality

Finally, undergirding the applicants' various challenges to the *vires* of the Regulations was the principle of legality. In particular, it was Lord Hoffmann's formulation in *R v Secretary of State for the Home Department Ex p Simms*¹⁷ (*Simms*) that was invoked and considered:

[T]he principle of legality means that Parliament must squarely confront what it is doing and accept the political cost. Fundamental rights cannot be overridden by general or ambiguous words ... In the absence of express language or necessary implication to the contrary, the courts therefore presume that even the most general words were intended to be subject to the basic rights of the individual.¹⁸

Unfortunately, the applicants seem to have invoked the principle of legality as a mere tack-on to buttress their other arguments. The impact of the principle and its role within the broader case received little attention. Instead, the applicants' argument contained little more than the assertion that 'the 1984 Act is subject to the principles in [*Simms*]', and thus, any ambiguity in its construction should be resolved in the applicants' favour.¹⁹

¹⁷ [2000] 2 AC 115 (HL).

¹⁸ *ibid* 131.

¹⁹ *Havers and Hoar* (n 9) [47].

2. *Decision of the Court of Appeal*

The Court of Appeal rejected the applicants' arguments challenging the *viros* of the Regulations, finding instead that 'the correct construction is that the Secretary of State did have power to enact the Regulations under the 1984 Act'.²⁰

A. The Principle of Legality

Although the principle of legality was but one of several subsidiary arguments advanced by the applicants, the Court of Appeal recognised the importance of addressing the issue.

However, the Court of Appeal sought to *avoid* applying the principle of legality by reframing the crux of the applicants' arguments. According to the Court, '[t]here can be no doubt ... that a justice of the peace has all of those powers'²¹—that is, to impose 'a special restriction or requirement'. Thus, in the view of the Court, the applicants' arguments merely turned on 'the relatively narrow [issue of construction] of whether the Secretary of State has power to impose such restrictions or requirements not only in relation to an individual or a group of persons but also in relation to the population generally in England'.²² Having downplayed the issue as not one concerning the *existence of a power*, but merely *the class of persons* against whom that power may be exercised, the Court concluded that the issue was 'not, on proper analysis, touched by the principle of legality'.²³

Moreover, the Court of Appeal held that, on the facts, the principle of legality was not triggered since 'it is not sufficient

²⁰ [2020] EWCA Civ 1605, [2021] 1 All ER 780 [78].

²¹ *ibid* [68].

²² *ibid*.

²³ *ibid*.

that there may be an interference with fundamental rights; what is required is that such rights would otherwise be “overridden”.²⁴ The Court noted that, in any case, the principle of legality was displaceable by the necessary implication that the Secretary of State had power to enact the Regulations under the 1984 Act.²⁵

B. The Scaling Up Argument

With the principle of legality purportedly out of the way, the Court of Appeal rejected the applicants’ scaling up argument.

Firstly, the Court of Appeal rejected the applicants’ submission that the general power conferred on the Secretary of State by Section 45C(1) is cut down by later provisions concerning ‘special restrictions or requirements’.²⁶ Support for this conclusion was gleaned from the provisions’ use of phrases such as ‘may in particular include’²⁷ and ‘include in particular’.²⁸ Moreover, the Court held that the applicants’ reliance on the statute’s use of the singular grammatical form in Section 45F foundered on the fact that the use of the plural is found elsewhere.²⁹ Accordingly, there is no contrary intention to displace the presumption that, in statute, words in the singular include the plural per Section 6(c) of the Interpretation Act 1978.³⁰

Secondly, the Court of Appeal rejected the submission that the Secretary of State’s power to impose ‘special restrictions or requirements’ is coterminous with the power of a justice of the

²⁴ *ibid* [67].

²⁵ *ibid* [65].

²⁶ PH(CD)A 1984 s 45C(4)(d).

²⁷ *ibid* s 45C(3).

²⁸ *ibid* s 45C(4).

²⁹ *ibid* s 45C(3)(c).

³⁰ [2020] EWCA Civ 1605, [2021] 1 All ER 780 [64].

peace. After all, the Court reasoned, ‘[i]f all that was required by way of a public health response was orders in respect of individuals or groups of persons, no doubt it would suffice to make an application to a justice of the peace’.³¹ Unless the Secretary of State’s power is more extensive than a justice of the peace’s, the 2008 Amendments would have been entirely redundant. Furthermore, the Court held that express exclusion of the making of certain orders from the scope of the Secretary of State’s power³²—despite them being available to a justice of the peace—suggests that the statutory regime already took into account the fact that the Secretary of State’s power was subject to fewer safeguards.

Thus, the Court of Appeal held that ‘Parliament intended the Secretary of State to be able to impose the other types of restrictions and requirements’ as could a justice of the peace, without the same constraints.³³

C. Necessary Implication

Ultimately—by way of dismantling the applicants’ scaling up argument—the Court of Appeal concluded that the Secretary of State had, by necessary implication, the power to enact the Regulations under the 1984 Act. Unpacking *precisely* how the Court arrived at this conclusion is important to our discussion in Part 3.

Having framed the issue of construction narrowly at the outset,³⁴ the Court of Appeal’s consideration of the necessary implication of the 1984 Act fixated on whether regulations

³¹ *ibid* [59].

³² PH(CD)A 1984 s 45D(3).

³³ [2020] EWCA Civ 1605, [2021] 1 All ER 780 [61].

³⁴ See text accompanying n 21-23.

enacted by the Secretary of State could affect the entire population in general.³⁵ As such, the Court did not feel the need to address the applicants' independent arguments challenging the *vires* of Regulations 6 and 8, but considered the *vires* of the Regulations all together—as if one in the same.

The reasoning employed by the Court of Appeal was as follows. Firstly, the Court cited *dicta* of Lady Hale from *R (Black) v Secretary of State for Justice (Black)*³⁶ to emphasise that a necessary implication is one that flows from the 'purpose, as well as the context, of the legislation'.³⁷ Secondly, the Court held that 'the purpose of the [2008 Amendments] clearly intended giving the relevant Minister the ability to make an effective public health response to a widespread epidemic'.³⁸ Thirdly, the Court held that '[i]f the power to make regulations were as limited as [the applicants' scaling up argument contended], it would not be effective in achieving that purpose'.³⁹ Therefore, the Court concluded that the 1984 Act conferred on the Secretary of State, by necessary implication, the power to enact the Regulations.

D. The Relevance of the 2004 Act

Having come to the 'clear conclusion' that the Regulations had been validly enacted under the 1984 Act, the Court of Appeal held that '[t]hat conclusion [was] not affected by the fact that the Secretary of State might have had power to make the regulations under the 2004 Act as well'.⁴⁰ That is despite the Court accepting

³⁵ [2020] EWCA Civ 1605, [2021] 1 All ER 780 [71], [78].

³⁶ [2017] UKSC 81, [2018] AC 215 [36].

³⁷ [2020] EWCA Civ 1605, [2021] 1 All ER 780 [70].

³⁸ *ibid* [71].

³⁹ *ibid* [59], [65].

⁴⁰ *ibid* [76].

that it ‘appear[s] to be correct’ that the Regulations were ‘of the kind that ... could have been made under the 2004 Act’.⁴¹

Considering the provisions of the 2004 Act, the Court of Appeal held that Parliament intended it as a last resort:

Under section 20(1) of the 2004 Act, Her Majesty may by Order in Council make emergency regulations if satisfied that the conditions in section 21 are satisfied ... One of the conditions ... is that (a) existing legislation cannot be relied upon without the risk of serious delay, (b) it is not possible without the risk of serious delay to ascertain whether the existing legislation can be relied upon, or (c) the existing legislation might be insufficiently effective ... the 2004 Act is an Act of last resort.⁴²

Thus, the Court of Appeal held that—as an Act of last resort—the 2004 Act could not logically preclude the Secretary of State from having a *valid choice* to enact the Regulations under the 1984 Act instead.⁴³

3. Legality in a Time of Emergency

Notwithstanding the backdrop of the Covid-19 pandemic in *Dolan*, it is well-established that in determining whether executive action falls within the four corners of statutory power, the

⁴¹ *ibid* [72].

⁴² *ibid* [72], [74].

⁴³ *ibid* [72]-[77].

exercise of construction ought to be the same in times of emergency as in times of normalcy.

In *HM Treasury v Ahmed*⁴⁴—a case involving the intersection between the principle of legality and ‘the kind of issue that led to Lord Atkin’s famously powerful protest in *Liversidge v Anderson*’—Lord Hope held that:

Even in the face of the threat of international terrorism, the safety of the people is not the supreme law. We must be just as careful to guard against unrestrained encroachments on personal liberty.⁴⁵

Lord Hope drew support from Lord Bingham’s extra-judicial remarks,⁴⁶ that ‘we are entitled to be proud that even in that extreme national emergency there was one voice—eloquent and courageous—which asserted older, nobler, more enduring values ... the role of the courts as guarantor of legality and individual right’.⁴⁷

This much is as true in the Covid-19 pandemic as ever.

As Jason Varuhas argues in an article published in the midst of the pandemic—unlike grounds of ‘substantive review’ which Courts tend to approach deferentially, the determination of the legal limits of power by ‘statutory interpretation is a matter quintessentially for courts, in respect of which courts exercise primary judgement and afford *no* deference’.⁴⁸ In a similar vein,

⁴⁴ [2010] UKSC 2, [2010] 2 AC 534.

⁴⁵ *ibid* [6].

⁴⁶ *ibid*.

⁴⁷ Lord Bingham, ‘The case of *Liversidge v Anderson*: The Rule of Law Amid the Clash of Arms’ (2009) 43 *The International Lawyer* 33, 38.

⁴⁸ Jason Varuhas, ‘The Principle of Legality’ (2020) 79(3) *CLJ* 578, 611. See also discussion in *Anisminic Ltd v Foreign Compensation Commission* [1969] 2 AC 147 (HL) 184, 207.

Lady Hale, in the 2019 Sir David Williams Lecture, cited the principle of legality as an area of public law that has mostly escaped the long reach of pragmatism.⁴⁹ However, not everyone is this optimistic. To David Dyzenhaus, the ‘compulsion of legality’⁵⁰ in times of emergency has caused many-a-Court to stretch and hollow out the concept of legality to the point of mere pretence.⁵¹

Dolan is a case that should give pause to the optimists. With respect, the Court of Appeal reached an executive-minded outcome only by unconvincing sleight of hand. The illusion involved two tricks: (1) circumventing the principle of legality and (2) smoke and mirrors in construing the Secretary of State’s power under the 1984 Act.

A. Circumventing the Principle of Legality

In *Dolan*, the Court of Appeal went to great lengths to avoid triggering the principle of legality. Unfortunately, its purported way of achieving this is problematic.

Firstly, the Court of Appeal’s assertion that the principle of legality was not engaged because the issue of construction did not concern the *existence of a power*—but merely the *class of persons*

⁴⁹ Lady Hale, ‘Principle and Pragmatism in Public Law’ (Sir David Williams Lecture, 18 October 2019) <<https://www.supremecourt.uk/docs/speech-191018.pdf>> accessed 10 April 2021, 15.

⁵⁰ ‘Compulsion of legality’ is a phrase coined by David Dyzenhaus to describe the compulsion felt by Courts to act in accordance with the law even in times of emergency. However, he argues that this often backfires as Courts defer excessively to the executive *in order to* lend judicial assistance to the emergency response.

⁵¹ David Dyzenhaus, ‘The Compulsion of Legality’ in Victor Ramraj (ed), *Emergencies and the Limits of Legality* (CUP 2008) 56.

against whom it may be exercised—is untenable. Whether an issue of construction concerns the existence of a power, or the class of persons subject to it, turns on little more than ingenuity of phrasing. This is because, fundamentally, the question of whether a *specific* power exists cannot possibly be divorced from the very questions identifying its scope and extent—who may exercise it, against whom may it be exercised, and subject to what procedural constraints?

Secondly, the Court of Appeal’s assertion that the principle of legality would *only* be engaged where the fundamental right in question would otherwise be ‘overridden’—mere ‘interference’ being insufficient—cannot be accepted. The Court seems to have reached this conclusion because of the particular usage of the word ‘overridden’ in Lord Hoffmann’s speech in *Simms*.⁵² However, to read *Simms* so constrictively is erroneous for two reasons.

In the first place, so constrictive a reading of *Simms* is inconsistent with recent cases of the highest authority in which the word ‘interference’ was preferred.⁵³ Indeed, if we were to accept the reasoning in *Dolan*, then in every case that the principle of legality is invoked, a Court must first determine whether the right in question is at risk of being ‘overridden’ or merely ‘interfered’ with. No authority was cited in *Dolan* for this, nor is it clear what the authority could be. On the contrary, this would be incompatible with the approach elucidated by Lady Hale in *J v Welsh Ministers*⁵⁴ concerning the legality of conditions imposed by

⁵² [2000] 2 AC 115 (HL) 131 (extracted above in text accompanying n 18).

⁵³ See *R (Unison) v Lord Chancellor* [2017] UKSC 51, [2020] AC 869 [79]; *J v Welsh Ministers* [2018] UKSC 66, [2020] AC 757 [24].

⁵⁴ [2018] UKSC 66, [2020] AC 757.

a community treatment order curtailing the liberty of a mentally ill patient:

We have to *start from the simple proposition* that to deprive a person of his liberty *is to interfere with a fundamental right* ... It is a fundamental principle of statutory construction that a power contained in general words is not to be construed so as to interfere with fundamental rights.⁵⁵

At no point did Lady Hale see the need to ask whether the conditions imposed on the patient amounted to an ‘overriding’ of, as opposed to a mere ‘interference’ with, his liberty. This suggests that the *true* controlling factor determining whether the principle of legality is triggered is not the *extent of interference*, but whether it is an interference *with a fundamental right or principle*.⁵⁶

Secondly, *even if Simms* could be read as authority for the proposition that the principle of legality attaches to the infringement and not the limitation of rights, Julian Rivers argues that this proposition rests on a conceptual confusion.⁵⁷ The argument that the principle of legality should be constricted in this way assumes that the principle exists ‘against a background of *conflict* between Parliament and the courts, with the courts grudgingly being forced by statute to countenance “unconstitutional” behaviour by the executive’.⁵⁸ However, as we will develop in Part 4, once we recognise that the principle of legality is in fact about *cooperation* between Parliament and the

⁵⁵ *ibid* [24] (emphasis added).

⁵⁶ For discussion of which rights and principles have been held to trigger the principle of legality, see Varuhas (n 48) 580-587.

⁵⁷ Julian Rivers, ‘Constitutional Rights and Statutory Limitations’ in Matthias Klatt (ed), *Institutionalized Reason: The Jurisprudence of Robert Alexy* (OUP 2012) 265.

⁵⁸ *ibid*.

Courts, it follows that the principle of legality applies to ‘ensure that limitations of individual rights in favour of other public interests gain sufficient consideration and legitimation by the democratically representative legislature’.⁵⁹

B. Smoke, Mirrors and a Necessary Implication

None of the preceding analysis would matter if the Court of Appeal was correct in finding that the Secretary of State had—by necessary implication—the power to enact the Regulations under the 1984 Act (the requisite necessary implication). If this was so, then any presumption which would have arisen by application of the principle of legality, would in any case have been displaced.

Unfortunately, however, the reasoning by which the Court purported to find the requisite necessary implication is problematic.

I. Putting the Cart before the Horse

In applying for leave to appeal to the Supreme Court, the applicants’ main complaint in relation to the Court of Appeal finding the requisite necessary implication was that the Court’s reasoning put the ‘cart before the horse’.⁶⁰ The applicants relied on the warning—by Lady Hale in *J v Welsh Ministers*—against ‘tak[ing] the assumed purpose [of legislation] ... and work[ing] back from that to imply powers ... which are simply not there’.⁶¹

In the context of *Dolan*, the applicants contended that:

⁵⁹ *ibid.*

⁶⁰ Philip Havers, Francis Hoar, and Wedlake Bell LLP, ‘Annex to Application for Permission’ (Published Court Document, 4 December 2020) [31].

⁶¹ [2018] UKSC 66, [2020] AC 757 [24].

The ‘cart’ in this case is the supposition that Parliament ‘must’ have intended to confer powers on Ministers to restrict persons other than those who ‘may be infected’ because those were the restrictions the Minister ... decided were necessary.⁶²

There is some merit to this complaint. It is notable that the Court of Appeal—despite citing Lady Hale’s judgement in *Black*—conveniently omitted Her Ladyship’s caution in the same case that ‘[i]n considering the intention of the legislation, it is not enough’ to say that its purpose is ‘the public good’, and that accordingly, the preferred meaning is correct because it is ‘more beneficial for the public’.⁶³

This distinction between a *necessary* and *efficacious* implication was arguably obliterated by the Court’s broad statement of the purpose of the 2008 Amendments—‘giving the relevant Minister the ability to *make an effective* public health response to a widespread epidemic’.⁶⁴ To an extent, the Court elided *necessary* and *efficacious* implications by asking whether its preferred meaning was *necessary for an efficacious* result.

At the same time, in our view, the applicants’ complaint points *less* to any definitive error in the Court of Appeal’s reasoning and *more* to an inherent difficulty in the exercise of statutory construction. This is particularly true in times of emergency, when the line between a *necessary* as opposed to *efficacious* implication will often be blurred beyond recognition.⁶⁵

⁶² Havers, Hoar, and Wedlake Bell LLP (n 60) [32].

⁶³ [2017] UKSC 81, [2018] AC 215 [36].

⁶⁴ [2020] EWCA Civ 1605, [2021] 1 All ER 780 [71] (emphasis added).

⁶⁵ Resort to looking at Hansard would likely not have been determinative in *Dolan*. Although the legislative background makes clear that the 2008 Amendments were intended to authorise *some*

II. *The Central Fallacy*

Far more problematic is the subtle yet fatal error which underpins the Court of Appeal's reasoning in finding the requisite necessary implication. At its crux, the Court appears to have fallaciously assumed that the necessary corollary of the failure of the scaling up argument was that the Regulations were validly enacted (the central fallacy). From this spawned a series of errors in the Court's reasoning.

III. *The Vires of Regulations 6 and 8*

One manifestation of the central fallacy was the Court of Appeal proceeding on the basis that all of the Regulations would stand or fall together, instead of taking the care required of it to consider whether *each* of the Regulations came within the power conferred on the Secretary of State.

Thus, the Court of Appeal entirely neglected the independent lines of argument challenging the *vires* of Regulations 6 and 8. This is deeply problematic. While it is true that if the applicants succeeded in their scaling up argument then *all* of the challenged Regulations would have been *ultra vires*, the converse is not. The mere fact that the scaling up argument foundered did not *necessarily* mean that each of the Regulations were validly enacted—dismantling the scaling up argument was necessary *but not* sufficient.

curtailment of individual liberties, it tells us little as to the *precise extent* of curtailment authorised. For contrasting views, see Sandhurst and Brandreth (n 10) 6; Jeff King, 'The Lockdown is Lawful' (*UK Constitutional Law Blog*, 1 April 2020) <<https://ukconstitutionallaw.org/2020/04/01/jeff-king-the-lockdown-is-lawful/>> accessed 10 April 2021.

The egregiousness of this error is underscored by the fact that Regulations 6 and 8 were clearly of a different and significantly more draconian character than the other Regulations. There was simply no reason for the Court to assume—without due consideration—that a single line of reasoning would apply to all the Regulations uniformly.

As illustration, consider the following.

On the Court of Appeal’s construction of the 1984 Act, the Secretary of State has the power to impose by regulations certain ‘restriction[s] or requirement[s] of the type which could be imposed by a justice of the peace’.⁶⁶ The most specific type of restriction that the Secretary of State was expressly authorised to impose was ‘restrictions on where P goes or with whom P has contact’, as specified in Section 45G(2)(j).

Thus, the question the Court of Appeal needed—but failed—to address was whether those words necessarily implied that the Secretary of State had the power to impose home confinement (which, without legal authority, would amount to the tort of false imprisonment).⁶⁷ Contrary to the Court’s judgement, this is a question to which the principle of legality *preeminently* applies. Even on the Court’s erroneous analysis, it is plain that the commission of a tort would amount to the ‘overriding’ of a fundamental common law right. If Section 45G(2)(j) does not contain the required necessary implication, then *a fortiori*, the Secretary of State would not be assisted by the more general provisions of Section 45C.⁶⁸

⁶⁶ [2020] EWCA Civ 1605, [2021] 1 All ER 780 [60].

⁶⁷ Text accompanying n 10-13.

⁶⁸ *R (Ingenious Media Holdings Plc) v Commissioner for HM Revenue and Customs* [2016] UKSC 54, [2016] 1 WLR 4164 [20].

However, the Court of Appeal's evasion left a 'significant question mark over whether section 45G(2)(j) can bear the weight that is placed upon it'.⁶⁹ After all, the words of Section 45G(2)(j) can encompass restrictions that either amount to, or fall short of, the commission of a tort.⁷⁰ A restriction 'not to go to x' is substantially different from a restriction to 'be confined *only* to x'. Even if the words of the provision can *literally* be read as encompassing both, per *R (Gedi) v Secretary of State for the Home Department*,⁷¹ such a provision will not be taken to authorise the imposition of home confinement by necessary implication. Accordingly, the principle of legality would not be displaced.

This is just one of multiple compelling lines of argument challenging the *vires* of Regulations 6 and 8 that were entirely overlooked as a result of the Court of Appeal's lax and broad-brush approach to statutory construction.

IV. Begging the Question and the Relevance of the 2004 Act

Another consequence of the central fallacy was the Court of Appeal's assumption that *because* the Secretary of State's power under the 1984 Act could not be as limited as the applicants contended, *therefore* it was as extensive as the Secretary of State contended.⁷² This blinkered approach to construction left the Court no meaningful room to consider the third possibility that *only another Act*—namely, the 2004 Act—conferred the requisite power on the Secretary of State to enact some, if not all, of the Regulations.

⁶⁹ Hickman, Dixon, and Jones (n 10) [34].

⁷⁰ *ibid.*

⁷¹ [2016] EWCA Civ 409, [2016] 4 WLR 93.

⁷² Text accompanying n 34-39.

Granted, the Court of Appeal's reasoning concerning the irrelevance of the 2004 Act is superficially attractive. *If* it is clear that the Regulations came within the scope of the 1984 Act, *and* the 2004 Act was intended by Parliament as a last resort, *then* why should the 2004 Act be at all relevant to the construction of the 1984 Act?

However, this is question-begging reasoning that adds nothing to the analysis. To say that the 2004 Act is not relevant because it is an Act of last resort is to simply *presuppose* that the 1984 Act is the Act of first resort. It does not respond to the applicants' contention that the *prior enactment* of the 2004 Act is *precisely* what makes the intended scope and effect of the 2008 Amendments unclear.⁷³

Instead, the Court offered little more than a glib assertion that 'the Secretary of State may well have had a choice of options and could have acted under' either the 2004 Act or the 1984 Act as amended in 2008.⁷⁴ Why this should be the case, however, is not obvious. Why should we readily accept that Parliament—having specified the appropriate level of scrutiny of emergency powers under the 2004 Act—subsequently conferred on the Secretary of State the power to *opt out* of that scrutiny for no apparent reason other than a choice of enabling legislation? Is this a question to which the principle of legality might apply?

We think, possibly. The principle of legality has not stood still since Lord Hoffmann's formulation of it in *Simms. R (Evans) v Attorney General*⁷⁵ (*Evans*) and *R (Privacy International) v Investigatory Powers Tribunal*⁷⁶ (*Privacy*) have extended the principle beyond

⁷³ Text accompanying n 16.

⁷⁴ [2020] EWCA Civ 1605, [2021] 1 All ER 780 [77].

⁷⁵ [2015] UKSC 21, [2015] AC 1787.

⁷⁶ [2019] UKSC 22, [2020] AC 491.

common law and human rights to encompass fundamental constitutional principles—albeit, limited in those cases to principles concerning the Court’s *own* institutional role within the constitutional order. The final step came in *R (Miller) v Prime Minister*⁷⁷ (*Miller 2*), which recognised the ability for the constitutional principle of Parliamentary accountability to serve as a hard-edged limit to executive action. Although *Miller 2* concerned the prerogative, a *very close* analogy exists between the principle of legality and what the Supreme Court was effectively doing in *Miller 2*. Indeed, Alison Young sees in the latter the application of a sister principle to the principle of legality.⁷⁸

How would this logic apply to the facts of *Dolan*? Prior to the 2008 Amendments, the Regulations would have had to be enacted under the 2004 Act. Instead, in the event, they were enacted under the 1984 Act as amended in 2008. As Keith Ewing argues, the safeguards built into the 1984 Act are so mild as to effectively permit ‘government by decree’ during the Covid-19 pandemic—especially given the mind-boggling pace of changes in regulations.⁷⁹ We argue that it would be *at most* a modest extension of the principle of legality to presume that Parliament did not intend—by the 2008 Amendments—to confer on the Secretary of State the power to evade Parliamentary accountability. This presumption would only be displaceable by

⁷⁷ [2019] UKSC 41, [2020] AC 373.

⁷⁸ Alison Young, ‘Prorogation, Politics and the Principle of Legality’ (*UK Constitutional Law Blog*, 13 September 2019) <<https://ukconstitutionallaw.org/2019/09/13/alison-young-prorogation-politics-and-the-principle-of-legality/>> accessed 10 April 2021.

⁷⁹ Ewing (n 16) 14-15; see also the discussion in *Dolan* at [29], [116]-[121].

‘concrete terms’ evidencing a ‘consistency of theme’⁸⁰ in the 1984 Act—things arguably absent given the serious ambiguities left by the 2008 Amendments.

In any case, the preceding discussion demonstrates that the Court of Appeal’s reasoning in finding the requisite necessary implication has little to stand on. More troublingly, when taken together, the Court’s judgement reveals an uncritical willingness to provide legal cover for executive action in a time of emergency.

*4. Legality in the Model of Legislative Accommodation*⁸¹

A. The Various Models of Emergency Powers

I. Interpretive Accommodation

Not everyone agrees that Courts must approach questions of the legality of executive action with the same stringency in times of emergency as in times of normalcy. On the contrary, Hanna Wilberg proposes responding to the exigencies of emergencies through transparent interpretive accommodation. Addressing the principle of legality in relation to the Regulations, she argues that:

⁸⁰ *B (A Minor) v Director of Public Prosecutions* [2000] AC 428, 473 (Lord Steyn).

⁸¹ For Part 4, we have borrowed the typology of constitutional models for emergency response—(i) interpretive accommodation, (ii) legislative accommodation, and (iii) extra-legality—from the work of Oren Gross and Fionnuala Ní Aoláin, *Law in Times of Crisis: Emergency Powers in Theory and Practice* (CUP 2006).

Everything may depend, then, on a balancing exercise: not just how important is the right and how severe the limit, but also how serious and how urgent was the emergency. The point I would emphasise is simply that we should not automatically assume that justified limits on liberty rights can never be authorised by general or ambiguous words.⁸²

The attractiveness of this approach, as Oren Gross and Fionnuala Ní Aoláin explain, lies in '[confronting] the inevitable by allowing it rather than by futilely resisting it'.⁸³ If Courts are going to reach pragmatic and deferential decisions in times of emergency *anyway*, then transparent recognition of that fact may obviate the need for judicial wizardry and allow for the direct evaluation of the particular balance struck in any given case.

In our view, however, there are serious difficulties and dangers in accepting interpretive accommodation as sound constitutional practice.

For a start, the method advocated by Wilberg problematically requires Courts to balance incommensurables—the demands of an emergency on one hand and individual rights on the other. The difficulty of doing this in an emergency context is underscored by the incredibly divisive dilemmas currently facing the ongoing Covid-19 response. As Martin McKee and David Stuckler's article demonstrates, determining the right

⁸² Hanna Wilberg, 'Lockdowns, the principle of legality, and reasonable limits on liberty' (*UK Constitutional Law Blog*, 23 July 2020) <<https://ukconstitutionallaw.org/2020/07/23/hanna-wilberg-lockdowns-the-principle-of-legality-and-reasonable-limits-on-liberty/>> accessed 10 April 2021. Wilberg argues this on the basis that the principle of legality contains a proportionality component—an argument which we address in Part 4.

⁸³ Gross and Ní Aoláin (n 81) 80.

balance to strike between (short and long -term) health, the economy and liberty is a multifaceted and cross-cutting question to which there is no easy answer.⁸⁴ Entrusting such determinations to the Courts is *highly* questionable since Courts lack the institutional expertise, evidentiary procedure and democratic legitimacy required for the task.

Moreover, as Frederick Schauer warns, this approach ‘runs the risk that the message of allowance will be taken as saying substantially more than it actually says, or allowing more than it actually allows’.⁸⁵ The threat of executive overreach left unchecked by an apologetic judiciary cannot be overstated. As Gross and Ní Aoláin point out, the belief that special judicial deference can be confined to the ‘exception’ is based on the false assumption that the separation between times of normalcy and times of emergency is clear-cut and defined. On the contrary, history shows that draconian measures adopted in times of emergency have the tendency to linger and infect the ordinary constitutional order.⁸⁶

II. Extra-Legality

Others have argued that the existence of an emergency threatening the very fabric and existence of society (note the indeterminacy of this descriptor—does the Covid-19 pandemic qualify?) justifies departing from the law on the logic of instrumentality. If, in such circumstances, the law fails to meet the demands of an emergency, is it not for the Benthamite ‘common

⁸⁴ Martin McKee and David Stuckler, ‘If the world fails to protect the economy, COVID-19 will damage health not just now but also in the future’ (2020) 26 *Nat Med* 640.

⁸⁵ Cited by Gross and Ní Aoláin (n 81) 80.

⁸⁶ Gross and Ní Aoláin (n 81) 159.

good? that the strictures of law be temporarily ignored? Consistent with this tradition, Albert Dicey famously argues that:

There are times of tumult ... when for the sake of legality itself the rules of law must be broken. The course which the government must then take is clear. The Ministry must break the law and trust for protection to an Act of Indemnity.⁸⁷

The most sophisticated version of this argument is that advanced by Gross and Ní Aoláin. They argue that—rather than have *ex ante* emergency legislation or have the Courts show deference to the executive in times of emergency—where the law runs out in the face of an emergency, the executive should simply and transparently declare that it will respond extra-legally and rely on *ex post facto* legitimisation of its actions. This, they argue, maintains a greater degree of separation between times of emergency and times of normalcy, preventing draconian measures from infecting the ordinary constitutional order.⁸⁸

While Gross and Ní Aoláin's argument raises interesting points—which unfortunately extend beyond the scope of this case comment—it suffices to say that we do not consider there to be any contradiction between what we envision the role of the Courts to be. This is because *any* model of extra-legality—unless it is willing to embrace a frankly unpalatable Schmittian conclusion⁸⁹—accepts that a *real prospect* of legal condemnation by the Courts is necessary to serve at least as a political restraint on

⁸⁷ Albert Dicey, *The Law of the Constitution* (Macmillan 1915) 272.

⁸⁸ Gross and Ní Aoláin (n 81) 159.

⁸⁹ Carl Schmitt was a legal scholar with fascist sympathies who argued that the law runs out in times of emergency, revealing the *true* sovereign as having absolute and unhindered freedom of action: see Carl Schmitt, *The Concept of the Political* (first published 1932, George Schwab, University of Chicago Press 1996).

the executive in times of emergency. Thus, Gross and Ní Aoláin agree with us that in times of emergency, Courts must take a ‘business-as-usual’ approach to statutory construction, leaving to Parliament the question of *ex post facto* legitimisation of otherwise *illegal* executive action.

Moreover, for the purposes of this case comment, we can remain agnostic as to Gross and Ní Aoláin’s argument that *ex ante* emergency legislation of any kind is undesirable. In our view, the most obvious difficulty with their normative model is that it is—as we will demonstrate—far-removed from constitutional doctrine and practice in the United Kingdom.

B. Legislative Accommodation

I. Democracy and the Principle of Legality

The dominant constitutional model for emergency response in the United Kingdom is legislative accommodation, whereby emergency powers are conferred on the executive by Parliament through an *ex ante* framework of ordinary legislation.

Nowhere is this more clearly evidenced than in the circumscription of the royal prerogative—that reservoir of executive power having been steadily displaced by statute. As Thomas Poole observes, ‘the main post-September 11th antiterrorism powers ... are all statute-based ... the emergency regimes that operated in both world wars were statutory in origin, as were the successive exceptional measures designed to deal with various emergency situations [since]’.⁹⁰ Thus, Lord Browne-

⁹⁰ Thomas Poole, ‘Constitutional Exceptionalism and the Common Law’ (2009) 7 *International Journal of Constitutional Law* 247, 253.

Wilkinson was able to say in *R v Secretary of State for the Home Department Ex p Fire Brigades Union* that:

The constitutional history of this country is the history of the prerogative powers of the Crown being made subject to the overriding powers of the democratically elected legislature as the sovereign body.⁹¹

As Eugen Ehrlich observes, the legal is inextricably bound up with the character of the society in which it operates.⁹² The model of legislative accommodation reflects the highly individualistic flavor of constitutionalism in the United Kingdom,⁹³ in which the executive cannot take unilateral action unsanctioned by the democratic mandate of Parliament. Such individualism is inherently skeptical of the executive's claim to know the 'common good', *unless* such knowledge is rooted in the individual experiences of the people as democratically represented. Thus, as Algernon Sidney put it, the executive 'has no other power than what the law allows'⁹⁴ since there is 'no other notion of wrong, than that it is a breach of the law which determines what is right'.⁹⁵ Accordingly, the constitutional role of the Courts is to guard against *any* executive action that usurps Parliament's democratic role by interfering with or modifying individual rights without the sanction of law—a responsibility that cannot be abdicated even in a time of emergency.⁹⁶

⁹¹ [1995] 2 AC 513 (HL) 552.

⁹² Eugen Ehrlich, *Fundamental Principles of the Sociology of Law* (HUP 1936) 390.

⁹³ Alan MacFarlane, 'The Origins of English Individualism: Some Surprises' (1978) 6 *Theory and Society* 255.

⁹⁴ Algernon Sidney, *Discourses Concerning Government* (first published 1698, Thomas West ed, Liberty Fund 1996) 222.

⁹⁵ *ibid* 284.

⁹⁶ Text accompanying n 57-59.

A common objection is that since emergencies are by definition unforeseeable, they therefore cannot be effectively legislated for *ex ante*. However, in our view, this argument attacks a strawman. Emergency legislation does not attempt to confer narrow powers in response to specific situations. Instead, they seek to subject the necessarily sweeping grants of executive power to the appropriate level of accountability determined *ex ante*, thus protecting Parliament's role in guiding the emergency response. As the background of *Dolan* illustrates, whether the Regulations were enacted under the 1984 Act or the 2004 Act was no mere technicality—it determined the scrutiny and control that Parliament had over extensive interference with individual rights.

The principles of democratic legitimacy thus require that the executive acts within its legal powers even in times of emergency. This necessarily commits us to the view (however inconvenient)⁹⁷ that a meaningful distinction exists between review concerning the *scope of statutory power* as opposed to the *manner of its exercise*. As Thomas Adams plainly states, ‘under the constitution, review for jurisdictional error features as a necessity’.⁹⁸ Statutory construction is quintessentially a matter for the Courts as the primary decision-maker and entails a robust willingness to nullify, as *ultra vires*, executive action exceeding the legal limits imposed by Parliament. As Lady Hale observes extra-judicially, ‘courts have been prepared to construe Acts of Parliament in the light of the principle of legality without a hint of deference or pragmatism’, and rightly so.⁹⁹

⁹⁷ Compare, for example, Rebecca Williams, ‘When is an error not an error? Reform of jurisdictional review of error of law and fact’ [2007] PL 793.

⁹⁸ Thomas Adams, ‘Ultra vires revisited’ [2018] PL 31, 39.

⁹⁹ Hale (n 49) 15.

II. Non-Deferential Judicial Posture in Times of Emergency

For the model of legislative accommodation to work, Courts *must* be willing to enforce the relevant statutory framework. As Laws LJ held in *R (Marchiori) v The Environment Agency*:

Democracy itself requires that all public power be lawfully conferred and exercised, and of this the courts are the surety. No matter how grave the policy issues involved, the courts will be alert to see that no use of power exceeds its proper constitutional bounds. There is no conflict between this and the fact that upon questions of national defence, the courts will recognise that they are in no position to set limits upon the lawful exercise of discretionary power in the name of reasonableness.¹⁰⁰

Reluctance by the Courts to do so defeats the *very purpose* of enacting emergency legislation and dangerously frustrates Parliament's intended safeguards for fundamental rights. Yet, despite judicial pronouncements and rhetoric to the contrary interspersed throughout the cases, *Dolan* is a reminder of the regrettable tendency for Courts to defer to the executive in times of emergency. This was so even though the Court of Appeal's decision in this case could not have undermined *any* ongoing emergency response, since the Regulations had already been repealed by then.¹⁰¹

¹⁰⁰ [2002] EWCA Civ 3, [2002] 1 WLUK 485, [40].

¹⁰¹ It was for this reason that the Court of Appeal considered the claim 'academic': [2020] EWCA Civ 1605, [2021] 1 All ER 780 [39]. The Regulations were amended on the 22 April, 13 May, 1 June and 13 June—before being repealed and replaced on the 3 July.

In the light of this, we agree with Dyzenhaus to the extent that *unless* Courts aspire and commit to discharging their constitutional responsibility in times of emergency, ‘the compulsion of legality¹⁰² results in the subversion of constitutionalism’ and the reduction of the model of legislative accommodation to mere pretence.¹⁰³ However, he goes too far in suggesting that the Courts’ constitutional responsibility is to, *as far as possible*, superimpose an ill-defined thick rule of law onto the construction of emergency legislation. To do so would give rise to the same problem of arrogating to the Courts the balancing of incommensurables.¹⁰⁴ As Poole incisively observes, Dyzenhaus does not explain why the values he believes constitute the thick rule of law ‘should outweigh (always? generally?) other, countervailing values, such as security or even national self-preservation’.¹⁰⁵

Instead, we argue that in times of emergency, Courts should aspire to take a business-as-usual approach to statutory construction and the principle of legality. Modifying Dyzenhaus’ argument, we argue that the self-consciousness imported by our aspirational model would serve as a necessary corrective to the tendency for Courts to defer to the executive in times of emergency. While, admittedly, some woolliness in application is unavoidable, we nonetheless think that this corrective would be effective and facilitate institutional cooperation within the model of legislative accommodation. Lord Atkin clearly did not think the notion that the law should ‘speak the same language’ a vacuous

¹⁰² See n 50.

¹⁰³ Dyzenhaus (n 51) 56.

¹⁰⁴ Text accompanying n 84.

¹⁰⁵ Poole (n 90) 265.

one,¹⁰⁶ and neither should we—it would likely have precluded the strained reasoning in *Dolan* itself.

Crucially, our argument upholds the principles of democratic legitimacy. By applying the principle of legality with the same *stringency* in times of emergency, Courts give due weight to the fundamental tenet that:

Parliament does not legislate in a vacuum. Parliament legislates for a European liberal democracy founded on the principles and traditions of the common law and the courts may approach legislation on this initial assumption.¹⁰⁷

At the same time, by applying the principle of legality *consistently*, the Courts provide Parliament with a predictable backdrop against which it can deliberately and explicitly design a framework of emergency powers. This approach shores up fundamental rights, respects Parliament's democratic right to strike the impossible balances needed to meet crises, and importantly, prevents the executive from exploiting hastily enacted legislation and flouting safeguards designed by Parliament.

III. Any Role for Proportionality?

One final point must be addressed. A premise of our argument is that Courts are justified in applying the principle of legality non-deferentially because the principle concerns the *scope* of executive power (as opposed to its *manner of exercise*) which is a matter quintessentially for the Courts. However, as Varuhas observes, there has been a tendency for Courts to disguise manner of

¹⁰⁶ See n 1.

¹⁰⁷ *R v Secretary of State for the Home Department Ex p Pierson* [1998] AC 539 (HL) 587 (Lord Steyn).

exercise review as applications of the principle of legality. Specifically, this has come in the form of Courts considering proportionality in delineating the scope of statutory power.¹⁰⁸

Thus, Timothy Endicott argues that while the Court of Appeal's reasoning in *Dolan* was problematic, its *outcome* could be justified on the basis that lockdown was 'necessary for the purposes for which Parliament conferred the power and ... the resulting detriment to liberty was not excessive in the light of that necessity'.¹⁰⁹ Respectfully, we disagree that proportionality has any role to play in the principle of legality.

In terms of authority, it is significant that proportionality has only ever been invoked in the legality inquiry *unidirectionally*—that is to justify a *more stringent*, as opposed to permissive, application.¹¹⁰ One might point to *Miller 2* as a counterexample,¹¹¹ however that case can be distinguished on the basis that it was the *absence of statutory language* which necessitated using reasonableness to delineate the scope of prerogative power.¹¹² Returning to the point, this observation suggests that the principle of legality is not *genuinely* concerned with proportionality at all. On the contrary, proportionality has been disguised within the principle of legality *precisely* to take advantage of the latter's 'strict scrutiny ... in cases where substantive review would not permit such scrutiny'.¹¹³ As

¹⁰⁸ See *R v Secretary of State for the Home Department, ex p Leech* [1994] QB 198; *R (Daly) v Secretary of State for the Home Department* [2001] UKHL 26, [2001] 2 AC 532 [15], [17]-[19] (Lord Bingham); *Simms* (n 17) 129 (Lord Steyn). For discussion, see Varuhas (n 48) 592-600, 606-614.

¹⁰⁹ Timothy Endicott, *Administrative Law* (5th edn, OUP 2021) 280.

¹¹⁰ Varuhas (n 48) 600.

¹¹¹ [2019] UKSC 41, [2020] AC 373 [50], [58] discussed in Wilberg (n 82).

¹¹² Varuhas (n 48) 599-600. In any case, *Miller 2* did not involve a structured proportionality assessment.

¹¹³ *ibid* 610.

Varuhas argues, this ‘surrogate form of ... substantive review ... raise[s] significant legitimacy concerns, not so far recognised or addressed by the Supreme Court’.¹¹⁴

Normatively, we think that the ingredients for rejecting the proportionality argument can be found in the preceding discussion. This argument fares no better in explaining why and how, in the context of an emergency, Courts should undertake the highly contentious balancing of incommensurables. Furthermore, as Endicott accepts, the proportionality argument does not work so well where the comparative is not between executive action or inaction in the face of an emergency, but instead—as was the case in *Dolan*—the *level of democratic accountability* that such executive action should be subject to.

Conceptually, once we recognise that the principle of legality is rooted in institutional cooperation within the model of legislative accommodation, it follows that the principle is not directly concerned with whether an interference with rights is proportionate or substantively justified. Rather, as Rivers argues:

The point is that the court refuses to judge of the substantive appropriateness of the balance which has been struck without the previous active consideration of the legitimate legislature. If there has been no such consideration, the court presumes that the question of balance has not been properly considered and finds the act unlawful.¹¹⁵

Essentially, democratic principles require the Courts to remit such questions to Parliament for further consideration. Of course, for a judge to do so in the midst of a pandemic takes

¹¹⁴ *ibid* 614. See also Adams (n 98) 42.

¹¹⁵ Rivers (n 57) 265.

courage. But until the Courts make clear that this is what they *must* and *will* do, we cannot begin designing stronger and more legitimate frameworks come the next test.

Conclusion

Dolan is a negative exemplar of undue judicial deference, which forces us to reconsider whether the law really speaks the same language in times of emergency as in times of normalcy. We argued that—as a matter of authority—the Court of Appeal could only reach the executive-minded outcome it did by a judicial sleight of hand in misconstruing the Secretary of State’s power under the 1984 Act. Our unpacking of the decision revealed a strained and unpersuasive attempt to circumvent and obfuscate the application of the principle of legality. Normatively, we argued that in times of emergency, Courts must aspire to resolve questions of the legality of executive action non-deferentially. Rejecting constitutional models of interpretive accommodation and extra-legality, we located the model of legislative accommodation within the United Kingdom’s liberal constitutional tradition. We argued that principles of democratic legitimacy require Courts to take a business-as-usual approach to statutory construction and the principle of legality even in times of emergency.

PRIVATE LAW ARTICLES

The Public Interest in Private Law: Re-Calibrating the Law of Tort

Juana De Leon^{*}

Abstract—The growing relevance of the public interest in the law of tort, particularly in the law of defamation and of nuisance, has led to the neglect of the private rights that it is charged to protect. This article argues that the law has in some instances too readily embraced the prioritisation of public over private interests without adequate scrutiny and without proper account of the fact that, in some areas, the private-public divide may not be as easily delineated. Tort law must be re-calibrated to reflect a more nuanced dynamic between these spheres: the law should emphasise private interests, ensuring that their true thrust is not curtailed by references to the public interest, without undermining the legitimacy of public interests altogether.

^{*} University College London. I am grateful to the Editorial Board of the OUULJ for their generous feedback on earlier drafts of this article. All errors are my own.

Introduction

The conflict between private and public interests, as well as their appropriate balance, has always been a contentious issue in the way the law of tort is understood and justified. This article will explore this conflict by focusing on the law of defamation and the law of nuisance, highlighting the different ways in which they demonstrate the private-public relationship. The substance of the private interests they are charged with protecting are distinct: defamation is a tort involving the *individual* and their reputation interest, whereas nuisance is a tort involving *land* and any interests that lie therein. There is generally thought to be a singular public interest that competes for recognition in the law of defamation (namely, freedom of speech); the interplay between private and public interests is therefore adversarial, with two interests in opposition to one another. This conflict is not as evident in the law of nuisance, given that the public interests in this area can vary, and are often plural and dynamic (including, for example, environmental protection or regional development). There is a degree of compromise and flexibility between these interests, in part because there is no pre-defined list of protected property rights, and also because of the concept of ‘reciprocity’, closely linked to the ‘reasonable user’ test,¹ which means that there is a ‘give and take’² between neighbouring parties that helps to draw the contours of their private interests.

Before investigating these relationships, a preliminary question ought to be asked: what is the aim of tort law? The ambition is not to rehearse the intricacies of each strand of scholarship, but simply to examine how some relate to the present discussion. Here, the inquiry is a normative one, seeking to

¹ *Cambridge Water Co v Eastern Counties Leather plc* [1994] 2 AC 264, 299.

² *ibid.*

resolve the conflict between tort as a mechanism of wider social justice, versus tort as a method of balancing private law rights. The former approach views the law of tort as a ‘social institution which exists not for its own sake but for the achievement of human goals and the performance of social functions’³ (the social aims approach). As Cane writes, ‘tort law rightly does not treat individuals as “islands unto themselves”, but as social creatures’.⁴ On this view, that tort law is a human enterprise, the law is ‘infected with the same tendency to pursue multifarious and potentially conflicting goals as are other forms of human purposive activity’.⁵ This means that courts almost *inevitably* consider broader interests extending beyond the parties to the litigation, thus characterising the courts as seekers of social aims rather than mere arbitrators of legal disputes.

By contrast, the school of ‘essentialism’ justifies the existence of tort law with the principle that people should not wrong others by their actions. As such, proponents of this view assert the oft-quoted slogan that ‘the purpose of private law is to be private law’.⁶ The claim made here is that the consideration of public interests in private law is ‘mistaken’⁷ and illegitimate; it is the individual, rather than any external factor(s), that acts as the principal moral agent. All responsibility, therefore, must flow from individual action. Weinrib presents challenges to the assumptions that underpin some variants of the social aims approach, in essence asserting that ‘recourse to independently valid goals implies the nonexistence of a distinctively legal mode of justification’.⁸ In other words, he argues that the social aims

³ Peter Cane, *The Anatomy of Tort Law* (Hart Publishing 1997) 205.

⁴ *ibid* 38.

⁵ *ibid* 224.

⁶ Ernest Weinrib, *The Idea of Private Law* (OUP 2013) 5.

⁷ *ibid*.

⁸ *ibid* 7.

approach only stands if it is accepted that tort law is insufficient to justify itself *internally* and cannot be understood from within. In opposition, he argues that, ‘far from being surrogates for the operation of independently justifiable collective purposes, [the fundamental concepts of private law] are the juridical markers of the immediate connection between the parties’.⁹

The view developed throughout this article lies between these approaches. Tort law, more specifically the law of defamation and the law of nuisance, currently aspires to the social aims approach and thus has increasingly carved out space for the accommodation of public interests. This is not a welcome development. The law, both in the form of legislation and in case law, has, in some instances, too readily accepted the prioritisation of public over private interests without adequate scrutiny and without proper account of the fact that, in some areas, the private-public divide may not be as easily delineated. In its attempt to implement the social aims approach, the law has diluted the private interests it was primarily tasked with protecting. This does not mean, however, that the pendulum ought to swing towards an entirely essentialist approach. An inherent difficulty with both approaches lies in the way they dichotomise private and public interests, characterising them as irreconcilable. This dichotomy can be seen in the Defamation Act 2013, which will be discussed in section 1, and in planning permission in private nuisance (particularly following the decision in *Coventry v Lawrence*¹⁰ (*Coventry*)), which will be discussed in section 2. On this basis, the view taken here focuses on *re-calibrating* the law of tort to reflect a more nuanced dynamic between private and public interests: the law should emphasise private interests, taking care to ensure that their true thrust is not stunted by references to the public interest

⁹ *ibid* 8.

¹⁰ [2014] UKSC 13.

(as has been the case in the adoption of the social aims approach) without invalidating the existence of public interests altogether (as under the essentialist approach).

1. Defamation

The Defamation Act 2013 (the Act) was passed to address the concern that the common law definition of defamation had a ‘chilling effect’¹¹ on freedom of expression. Section 1 (the serious harm requirement) has succeeded in protecting freedom of expression, but it has done so at the expense of a robust protection of reputation. The introduction of the requirement poses difficulties for corporate claimants (now needing to demonstrate financial loss for this threshold) and overlooks the value of corporate reputation by failing to account for the idea that the neglect of private interests may sacrifice the protection of public interests that are closely interlinked with them. Contrary to the notion that freedom of speech is the only public interest at play, the discussion will show that, although corporate reputation is a private interest, it *also* contains elements of the public interest (namely in economic mobility), and in failing to protect corporate reputation, the law fails to adequately protect both the private and the public interest. Private and public interests cannot simply be disaggregated from one another, and thus they cannot be deemed to exist in a see-saw relationship whereby stronger protection of the former necessarily means lesser protection of the latter, and

¹¹ Anthony Lester, ‘These Disgraceful Libel Laws Must Be Torn Up’ *The Times* (London, 15 March 2011) <<https://www.thetimes.co.uk/article/these-disgraceful-libel-laws-must-be-torn-up-698t8zpdk9w>> accessed 28 September 2020.

vice versa. Public interests can be plural and dynamic: safeguarding one set of interests does not automatically ensure holistic protection of all public interests.

A. Serious Harm and Financial Loss

One of the Act's first priorities was 'to ensure that the law is reformed so that trivial and unfounded actions for defamation do not succeed'.¹² As such, section 1(1) of the Act 'raises the bar for a statement to be defamatory by [requiring] that it must have caused or be likely to cause serious harm'.¹³ It was confirmed in *Lachaux v Independent Print Ltd*¹⁴ (*Lachaux*) that this new threshold not only superseded its common law predecessors, but also requires reference to the actual or likely impact of the words as opposed merely to their inherent meaning. The need to consider the actual or likely impact of an alleged defamatory statement means that the previous common law presumption of damage to the claimant's reputation is no longer tenable.¹⁵ This presumption meant that a claimant could establish the defamatory nature of a statement by reference solely to the words used and could therefore recover damages without establishing actual or likely harm to their reputation. It was thought to be irrebuttable in practice,¹⁶ thus enhancing the protection afforded to reputation. Although counsel for the claimant tried to argue that the phrase 'likely to cause' in section 1 meant that the presumption survived the Act, the court in *Lachaux* made clear that this provision

¹² HL Deb 9 October 2012, vol 739, col 934.

¹³ *ibid.*

¹⁴ [2019] UKSC 27.

¹⁵ *ibid* [17].

¹⁶ *Multigroup Bulgaria Holding AD v Oxford Analytica Ltd* [2001] EMLR 28; *Jameel (Yousef) v Dow Jones Co Inc* [2005] QB 946.

‘introduce[d] a new threshold of serious harm which did not previously exist’.¹⁷

This is a significant change in the balance struck between private and public law considerations. Section 1(1), by removing the presumption and substantively raising the threshold for a statement to be defamatory, ensures stronger protection for the public interest in freedom of expression. For the purposes of this threshold, for-profit bodies under subsection (2) must now show actual or likely ‘financial loss’. Below it is argued that this requirement, particularly when read with the newly established substantive threshold, makes it difficult for corporate claimants wishing to vindicate their reputations under the Act. As will be discussed in Section B, the financial loss requirement reflects an incomplete picture of the nature of corporate reputation (only capturing reputation as property) and does not account for pecuniary losses that may flow from damage to reputation as honour. These protracted losses for companies in turn trigger concerns about the interests of communities that depend on the financial success of these corporations. Section C canvasses an alternative model to corporate defamation claims wherein financial loss is removed as a substantive qualifier in the finding a defamatory statement, and instead, a multi-factor version of the *Lachaux* paradigm is introduced.

B. The Nature of Corporate Reputation

Post has argued that there are three manifestations of reputation in the law of defamation: property, honour, and dignity.¹⁸ Lord Hoffmann has emphasised that a company’s standing to sue in defamation is rooted in its reputation as property, as ‘a

¹⁷ *Lachaux* (n 14) [13].

¹⁸ Robert Post, ‘The Social Foundations of Defamation Law: Reputation and the Constitution’ (1986) 74 Calif L Rev 691, 693.

commercial company has no soul and its reputation is no more than a commercial asset, something attached to its trading name which brings in customers'.¹⁹ The introduction of the financial loss requirement was intended to reflect this sentiment, while addressing the longstanding concern that 'Large corporations can and do use their financial muscle to stifle legitimate debate by threatening to sue their financially weaker critics.'²⁰ That financial loss must now be evidenced is reflective of the *harm*-orientation of reputation as property: any damage done to reputation can be shown through loss that flows as a *consequence* of defamation.

According to Post, although dignity and honour are generally thought to be synonymous, they serve analytically distinct purposes in the understanding of defamation law. Reputation as dignity is concerned with the individual's interest in being included within their community (infringement upon such an interest would lead that individual to be stigmatised and ostracised).²¹ This interest is rooted *within* the individual and is therefore inapplicable to companies. Conversely, reputation as honour is concerned with the normative characteristics of a particular social status and the corresponding 'regard and estimation that society accords to that role'.²² In contrast to the *harm*-orientation of reputation as property, reputation as honour is *wrong*-oriented. The wrong perpetrated is socially rooted, leading to a disruption of 'shared social perceptions that transcend the behavior of particular persons'.²³ This manifestation of reputation was reflected in the common law presumption of damage, which focused on a

¹⁹ *Jameel v Wall Street Journal Sprl* [2007] 1 AC 359 [91].

²⁰ HL Deb, 9 October 2012, vol 739, col 966.

²¹ Post (n 18) 711.

²² *ibid* 700.

²³ *ibid* 702.

“noncompensatory” end²⁴ of vindicating and reaffirming the claimant’s social status even in the absence of evidence indicating harm. Unlike reputation as dignity, reputation as honour is *externally* and *relationally* determined. It can serve as an analytical tool to show a more holistic picture of the losses that can be suffered by companies when their reputation is damaged, particularly in light of contemporary business practices. As Chan highlights when discussing reputation as honour, ‘Such reputation may not be ‘earned’ in the ‘marketplace’ proper for the purposes of attracting customers, but the good name of the organisation (as an advocate for social justice and compassion for the disadvantaged people, for instance) might have been built upon the work and efforts of the founders, employees or officers within the organisation.’²⁵ These efforts are now commonplace given the growing importance that many companies place on corporate social responsibility (CSR) practices.²⁶ This will often involve not-for-profit activities, the aims of which are to establish a positive image or brand for a company. For example, Apple has a reputation as regards the quality of its products (reputation as property) but also has, through its CSR activities, established itself as an active supporter of LGBTQ+ rights²⁷ (reputation as honour). Similarly, Ben & Jerry’s has a reputation as regards its ice cream (reputation as property) but has also cemented itself as

²⁴ *ibid* 706.

²⁵ Gary Chan, ‘Corporate Defamation: Reputation, Rights and Remedies’ (2013) 33 *Legal Studies* 264, 268.

²⁶ Peter Coe, ‘The Value of Corporate Reputation and the Defamation Act 2013: a Brave New World or Road to Ruin?’ (2013) 18(4) *Comms L* 113, 114.

²⁷ David Graham, ‘The Business Backlash to North Carolina’s LGBT Law’ (*The Atlantic*, 25 March 2016) <<https://www.theatlantic.com/politics/archive/2016/03/the-backlash-to-north-carolinas-lgbt-non-discrimination-ban/475500/>> accessed 16 December 2020.

committed to social, environmental, and political issues through its continued track record of corporate activism (reputation as honour).²⁸ This manifestation is particularly apparent in instances where company reputation is tightly bound up in the image of their founders, such that ‘poor judgment and unethical behavior on the part of company founders directly impacts positive perception of a company’.²⁹ A paradigmatic illustration of this is the calls to boycott In-N-Out Burgers’ following their donation to the Republican party, pursuant to the personal political beliefs of its founder and executives.³⁰ Again, such reputational damage disrupts ‘shared social perceptions’ of the company, which can likely lead not only to a decline in consumer support, but perhaps also in the ability to recruit employees or secure lucrative partnerships for the company.

There is a common objection to this characterisation of corporate reputation, namely that companies, although they are *legal* persons, are not imbued with the same substantive rights inherent in *natural* persons and are therefore undeserving of the same protection endowed by the courts. Connected to this is the

²⁸ Tal Donohue and Georgia Rhodes-Bell, ‘A Rocky Road for Brands: What Ben & Jerry’s Teaches Us About Corporate Activism, Purpose and Reputation’ (*Infinite Global*, 20 August 2020) <<https://infiniteglobal.com/infinite-brief/a-rocky-road-for-brands-what-ben-jerrys-teaches-us-about-corporate-activism-purpose-and-reputation-2/>> accessed 16 December 2020.

²⁹ Peter Daisyme, ‘Consider How Your Actions As A Founder Impact Your Brand And Business Reputation’ (*Forbes*, 13 November 2018) <<https://www.forbes.com/sites/theyec/2018/11/13/consider-how-your-actions-as-a-founder-impact-your-brand-and-business-reputation/?sh=6c9b37e3680b>> accessed 15 April 2021.

³⁰ Jeff Daniels, ‘In-N-Out Burger’s \$25,000 donation to California GOP brings call for boycott from Democrats’ (*CNBC*, 30 August 2018) <<https://www.cnn.com/2018/08/30/in-n-out-burger-faces-boycott-for-california-gop-donation.html>> accessed 15 April 2021.

concern that these CSR practices are merely ancillary to a company's identity or otherwise form part of a greater profit-making strategy that ultimately links back to their reputation as property. Although these concerns are not without merit, particularly given the growing concerns of companies wielding too much power, there is evidence to show that modern business practices are beginning to prioritise reputation best conceived as honour, in some instances above reputation as property, thus warranting protection in its own right. As explained by Post, the individual 'claims a right to [honour] by virtue of the status with which society endows his social role.'³¹ In turn, 'society expects him to aspire to "personify" these attributes and to make them part of his personal honour.'³² Social status in this sense sets the expectations *by* and *through* which the individual can navigate and elect their membership or non-membership to that status. This idea of social status is being prioritised by a new breed of businesses called 'B corporations'. B corporation status requires certification, which involves an assessment of a company's ability to 'create value for non-shareholding stakeholders, such as their employees, the local community, and the environment.'³³ They are *legally* required to consider these factors, and must therefore ensure that their organisational structure is not exclusively focused on profit maximisation.³⁴ Electing to undergo certification, or aspiring towards certification, 'is a way to publicly claim an identity as an organization interested in both shareholder

³¹ Post (n 18) 700.

³² *ibid.*

³³ Suntae Kim et al., 'Why Companies Are Becoming B Corporations' (*Harvard Business Review*, 17 June 2016) <<https://hbr.org/2016/06/why-companies-are-becoming-b-corporations>> accessed 27 March 2021.

³⁴ Isabella Fox, 'A guide to the UK B Corporation movement' (*PwC*) <<https://www.pwc.co.uk/industries/retail-consumer/insights/b-corp.html>> accessed 27 March 2021.

and stakeholder success'.³⁵ In fact, there is evidence to suggest that companies choose to become B corporations in order to distinguish themselves from profit-driven CSR practices 'in the midst of a "greenwash" revolution'³⁶ (whereby companies make unsubstantiated claims to deceive consumers into believing the ethicality and sustainability of their products or practices³⁷), which in turn allows individuals to differentiate between traditional firms and 'genuine, authentic advocates of stakeholder benefits'.³⁸ This demonstrates a somewhat reflexive relationship, similar to the process described by Post whereby companies behave in a way that responds to, as well as claims an identity that aligns with, the normative expectations set by consumers to establish and maintain a particular social status. Companies are now empowered to act in a way that cannot fully be accounted for by reference to profitability (reputation as property), instead better viewed as a commitment to the society in which they operate (reputation as honour).

When the social status garnered from this commitment is tarnished, the damage done to the company will not always *immediately* translate to financial loss. Such honour-damage can lead to an inability to recruit talented employees, to obtain property and facilities, or in the case of journalism, to gain access to individuals and sources for the delivery of timely and accurate news articles. These non-financial losses go to the heart of a company's ability to run their business, 'impair the purpose and

³⁵ Kim et al. (n 33).

³⁶ *ibid.*

³⁷ Leyla Acaroglu, 'What is Greenwashing? How to Spot It and Stop it' (*Disruptive Design*, 8 July 2019) <<https://medium.com/disruptive-design/what-is-greenwashing-how-to-spot-it-and-stop-it-c44f3d130d5>> accessed 15 April 2021.

³⁸ Kim et al. (n 33).

object of its establishment and maintenance’,³⁹ and can therefore be more protracted than, for example, a drop in share price immediately following a defamatory statement about the quality of a company’s products (reputation as property). Admittedly, these honour-losses may *ultimately* lead to financial losses. However, there is no guarantee that such losses, now further down the causation chain, can be linked back to the defamatory statement. The point to be stressed is that using financial loss as a prerequisite to the finding of whether a statement is defamatory only reflects reputation as property and does not capture the different *types* of losses that can be sustained by damage to reputation as honour. Again, the former is *harm*-oriented while the latter is *wrong*-oriented. To use financial loss as a qualifier to a company’s defamation claim is to collapse these orientations, which leads to the conclusion that ‘defamation does not protect reputation but the ‘temporal’ consequences of its violation, that is, those which can be valued in money’.⁴⁰

C. Corporate Reputation and Community Interests

Though deemed a private entitlement, corporate reputation has wider implications for the social interests and economic mobility of communities. As expressed in *Steel and Morris v United Kingdom*, ‘Although large public companies laid themselves open to close scrutiny of their acts and the limits of acceptable criticism were wider in the case of such companies ... there was a competing public interest in protecting the commercial success and viability of companies, not only for the benefit of shareholders and

³⁹ Chan (n 25) 270.

⁴⁰ Erik Descheemaeker, ‘Three Errors in the Defamation Act 2013’ (2015) 6(1) JETL 24, 30.

employees but also for the wider economic good'.⁴¹ This is not to say that corporations always *act* in the public interest, but it nevertheless acknowledges that their reputations have *ramifications* for the public interest in terms of economic mobility. This proposition is supported by the fallout of the Volkswagen emissions scandal; although the scandal did not involve a claim of defamation, it nevertheless illustrates the impact of damage to corporate—as well as collective—reputation on the economy.

Volkswagen is one of Germany's largest employers, creating more than 270,000 jobs in the country and even more working for suppliers.⁴² In 2014, approximately 775,000 people worked in the automobile sector, constituting almost two percent of the entire national workforce.⁴³ Moreover, the car industry is Germany's biggest exporter—an industry whose reputation relies heavily on the 'Made in Germany' brand to gain competitive edge in the market.⁴⁴ Volkswagen commercials reinforce this brand by ending with the slogan 'Das Auto' ('The Car'), subliminally messaging to consumers that their cars embody the model standard of the automobile industry. The emissions scandal, however, turned this brand into a liability, not only for Volkswagen, but also for German car companies.⁴⁵ Other German car companies, those who similarly relied on the 'Made

⁴¹ [2005] 41 EHRR 22 [94].

⁴² Carson Jones, 'Volkswagen and Its Effect on the Global Economy' (*globalEDGE*, 2 October 2015) <<https://globaledge.msu.edu/blog/post/31014/volkswagen-and-its-effect-on-the-german-economy>> accessed 16 December 2020.

⁴³ *ibid.*

⁴⁴ Dimitrije Ruzic, 'How the Volkswagen Scandal Turned "Made in Germany" Into a Liability' (*Insead Knowledge*, 16 August 2019) <<https://knowledge.insead.edu/economics-finance/how-the-volkswagen-scandal-turned-made-in-germany-into-a-liability-12196>> accessed 16 December 2020.

⁴⁵ *ibid.*

in Germany' brand, suffered from the scandal; this shift was described as a 'spillover effect as arising from collective reputation'.⁴⁶ The losses totalled approximately \$26.5 billion, following a decline in sales across the sector⁴⁷ and created well-founded concerns about the consequences on the German economy.⁴⁸

Admittedly, this is an extreme example and perhaps its gravity will not always feature in defamation claims. Nevertheless, it demonstrates a clear link between corporate reputation and economic mobility: companies and their reputations do *not* exist in a vacuum. They cannot always be disentangled from one another, nor can they be completely divorced from the public interest. Indeed, a company structure is not monolithic. Damage done to its reputation, even for those that are not corporate juggernauts, will have a negative impact upon its employees, its suppliers, and the wider community within which it operates.⁴⁹ As such, 'Locating reputation in a purely property-based dimension would signal that companies need not seriously treat externalities of their decisions as corporate obligations, since their corresponding reputational rights are not seriously protected.'⁵⁰

⁴⁶ Ruediger Bachmann et al., 'Firms and Collective Reputation: A Study of the Volkswagen Emissions Scandal' (2019) National Bureau of Economic Research Working Paper 26117, 30 <<https://www.nber.org/papers/w26117>> accessed 16 December 2020.

⁴⁷ *ibid* 3.

⁴⁸ 'Emissions Scandal is a Risk to Germany Economy: MBinistry' (*Reuters*, 20 August 2017) <<https://www.reuters.com/article/us-germany-economy-idUSKCN1B00U5>> accessed 16 December 2020.

⁴⁹ Alastair Mullis and Andrew Scott, 'Lord Lester's Defamation Bill 2010: a distorted view of the public interest?' (2011) 16(1) *Comms L* 6, 14-5.

⁵⁰ Cara Gao, 'Serious Harm to Corporate Reputation – More Than Just Money' (2018) 7 *OUULJ* 10, 27.

The issue highlighted here is the Act's misplaced binarization of private and public interests through the financial loss requirement. This is not to diminish the importance of the public interest in freedom of expression, but to demonstrate that casting this area as a stark conflict of private and public interests is misguided: private entitlements can be closely bound up with, and may therefore necessarily effect, public interests.

In light of the lack of clarity surrounding the private-public divide in this area, the law of defamation should re-centralise its emphasis on reputation. This could be done by removing the financial loss requirement *as a prerequisite* for the purposes of finding whether a statement is defamatory. As has been discussed above, corporations may suffer damage to reputation from defamatory statements in a way that does not manifest in just direct financial loss and 'private interest' damage. As such, the financial loss prerequisite should be replaced with a multi-factorial enquiry. In doing so, it is not suggested that financial loss should be removed entirely from consideration, but merely that it should stop operating as the sole yardstick against which all harm is quantified. Financial loss may be incorporated as a *factor*, and indeed a strong one, but other factors may include evidence of: a decline in job applications submitted to the company (as compared to previous years); an increased inability to obtain facilities, suppliers, permits or other instruments required to promote the efficacy of the business; a decrease in advertising opportunities etc. This is by no means an exhaustive list, but is indicative of the manifestations of reputational damage that ought not to be neglected by the law of defamation. Any issues of causation arising out of the difficulties in proving financial consequences occurring further down the causation chain would be alleviated, at least in part, as claimants would no longer need to prove that a statement caused actual or likely financial harm, but rather can paint a more holistic image of the

nature and extent of damage suffered. If it can be shown, for example, that a longstanding advertising partnership has been lost due to a defamatory statement, this can contribute to the picture of loss suffered by the for-profit body. These factors should have cumulative effect, but the relative weight attached to each factor may be decided on a case-by-case basis. For those concerned about barring trivial claims, ‘That trivial injuries ought not to get their day in court is naturally an ancient idea that is not necessarily objectionable ... but this is normally dealt with through the operation of rules that are *external* to the cause of action being found, for instance, in the rules of civil procedure.’⁵¹ Finally, all of these suggested changes should not undermine the efforts made to secure freedom of expression under the Act given that, post-*Lachaux*, the common law presumption of damage has not survived, and harm must still meet the new substantive threshold of seriousness under section 1(1).

2. Nuisance

Private nuisance is a tort against land that hinges on the ‘principle of reasonable user’,⁵² characterised as ‘the principle of give and take as between neighbouring occupiers of land’⁵³ and ‘the relevant control mechanism’⁵⁴ through which liability is determined. As mentioned in the introduction, this element of reciprocity means that, instead of a list of *pre-existing* rights sought to be protected, it is the relationship between the property rights

⁵¹ Descheemaeker (n 40) 27.

⁵² *Cambridge Water* (n 1) 299.

⁵³ *ibid.*

⁵⁴ *ibid* 300.

of neighbouring parties that helps to draw the contours of their private interests: it involves a balancing exercise of the interests held by both the claimant and defendant, consisting of several factors, such as the nature and the extent of the damage, the locality of the area,⁵⁵ the claimant's use of the land,⁵⁶ and the motive of the defendant.⁵⁷

The following discussion focuses on planning permission and its role within this exercise. Planning permission has proven to be an inappropriate vehicle for the expression of the public interest and is therefore an inadequate justification for the curtailment of private property rights when carrying out this balancing act. The Supreme Court in *Coventry*, however, seems to have placed greater emphasis on this purported expression of the public interest, particularly by highlighting the role that planning permission plays when determining the grant of an injunction. This is once again demonstrative of the courts' adoption of the social aims approach. With respect, this increased emphasis not only mischaracterises the power that ought to be attributed to administrative decisions, but also signals a departure from how the tort of nuisance is doctrinally conceived and loses sight of proprietary rights that private nuisance is meant to protect.

A. Coventry v Lawrence

The claimants in *Coventry*, who complained about the noise from the defendants' motor racing activities, were successful in establishing the defendants' liability in nuisance, despite the activities having the benefit of planning permission. The Supreme Court overturned the finding of the Court of Appeal that the grant of planning permission may change the character of locality,

⁵⁵ *St Helen's Smelting Co v Tipping* [1865] 11 HL Cas 642.

⁵⁶ *Network Rail v Morris* [2004] EWCA 172.

⁵⁷ *Hollywood Silver Fox Farm Ltd v Emmett* [1936] 2 KB 468.

holding that planning permission is ‘normally of no assistance to the defendant’⁵⁸ in the assessment of liability in private nuisance. It may provide a starting point for the assessment of reasonable use, but in many cases will be ‘of little, or even no, evidential value’.⁵⁹ However, the speeches in *Coventry* seem to share a common assumption, albeit of varying degrees, that the planning system is an adequate channel for the expression of the public interest. Planning permission may be indicative of an activity being of benefit to the public and, as such, may be relevant to the question of remedies (*how* the tort is to be redressed) as opposed to liability (*whether* a tort was committed).

The speeches in the judgment indicate a shift towards the liberalisation of damages in lieu of an award for an injunction. Perhaps most radically, Lord Sumption goes as far as to propose somewhat of a presumption against the grant of injunctions whenever an activity has the benefit of planning permission.⁶⁰ Lord Mance and Lord Carnwath disagreed with this proposition, with Lord Mance arguing that this afforded too much significance to planning permission and public benefit, particularly in light of the fact that the right to enjoy one’s home without disturbance is valued by many for reasons almost entirely independent of money.⁶¹ Lord Clarke thought that the existence of planning permission may be relevant to the question of the appropriate remedy, but His Lordship declined to lay down any precise principles to be followed in the future.⁶² Lord Carnwath acknowledged the tension between planning law and the law of nuisance, given that the former exists to promote the public interest, whereas the latter seeks to protect the private rights of

⁵⁸ *Coventry* (n 10) [94].

⁵⁹ *ibid* [96].

⁶⁰ *ibid* [161].

⁶¹ *ibid* [167]-[168] (Lord Mance), [246] (Lord Carnwath).

⁶² *ibid* [171]-[173].

particular individuals.⁶³ His Lordship agreed that emphasis ought to be placed upon private law rights at the liability stage of the enquiry, but the public interest may be relevant to the remedial question.⁶⁴ He also expressed his approval of the view that the continued strength of private nuisance in a regulatory state ‘probably depends on a flexible approach to remedies’.⁶⁵

Lord Neuberger, giving the leading judgment, attempted to homogenise the varying approaches, concluding that injunctions would generally remain the primary remedy while also providing a revision of the *Shelfer v City of London Electric Lighting Co*⁶⁶ (*Shelfer*) guidelines. *Shelfer* held that damages would be awarded only in the most limited circumstances, essentially amounting to a four-staged test that swung heavily in favour of injunctions: (1) if the injury to the claimant’s rights is small; and (2) is capable of being estimated in money; and (3) is adequately compensable by a small money payment; and (4) the case is one in which it would be oppressive to the defendant to grant an injunction.⁶⁷ There was a general sentiment within Their Lordships’ judgments that *Shelfer* was out of date and, throughout the years, was wielded far more often than the court deemed appropriate. Lord Neuberger clarified that fulfilling the *Shelfer* criteria will not automatically mean that an injunction should be granted, adding that their application should not be ‘a fetter on the exercise of the court’s discretion’.⁶⁸ The public interest, although never substantively defined in the judgment, was

⁶³ *ibid* [193].

⁶⁴ *ibid* [222].

⁶⁵ Maria Lee, ‘Tort Law and Regulation: Planning and Nuisance’ (2011) 8 JPL 986, 990.

⁶⁶ [1895] 1 Ch 287.

⁶⁷ *ibid* 322.

⁶⁸ *Coventry* (n 10) [120].

deemed as a matter of law ‘to be a relevant factor’⁶⁹ in this calculation. Despite the differing approaches taken by Their Lordships on this issue, what is nevertheless clear is a common accommodation of the public interest in the form of planning permission.

The readiness to incorporate planning permission into the question of remedies stems from a belief that this will strike the correct balance between private property rights and any public interest that may arise through the permission process. If the court were to consider planning permission at the liability stage, potentially allowing it to determine the finding of a nuisance, it would likely be vulnerable to the criticism that the public interest is being used to curtail the *assertion* of private property rights and thus the important recognition thereof. However, when local authority approval is shifted to the remedial question, this objection arguably loses its vigour; the claimant will have already had their property rights *recognised* and the question simply falls to be asked about how they are to be compensated. As will be explained, this distinction is artificial. Injunctions in private nuisance give meaning to the individual’s property rights, and the frequency with which they are awarded ought not to be limited by references to the (often unclear) public interests contained within planning permission.

B. The Relaxation of Shelfer and the Planning Process

The primary remedy in nuisance is an injunction for the very reason that it reflects and protects one’s proprietary right to enjoy one’s land without interference. By reserving for the courts a sizeable discretion over the award of an injunction, and by

⁶⁹ *ibid* [124].

effectively downgrading the *Shelfer* criteria, the judgment in *Coventry* overlooks the proprietary nature of the tort of nuisance. Admittedly, the concerns expressed by Their Lordships in *Coventry* are not unfounded, namely that previous courts have too readily accepted the grant of an injunction before adequately considering whether damages could be appropriate. Indeed, as highlighted by the court, the current landscape involves stronger market forces and higher degrees of regulatory control that reflect collective, but nevertheless legitimate, aspirations for development;⁷⁰ it is important that these aspirations are not undermined and therefore ought not to be neglected when drawing the boundaries of private rights. However, it should be remembered that nuisance is a tort involving *land*, governed by ‘property rules’, rather than ‘liability rules’.⁷¹ The position in *Coventry* conflicts with the doctrinal orthodoxy that ‘a person by committing a wrongful act ... is not thereby entitled to ask the Court to sanction his doing so by purchasing his neighbour’s rights’.⁷² The very essence of property rights is that they are distinct from ‘mere expressions of a contractual bargain’,⁷³ under which a defendant can avoid liability by paying a suitable price. A hallmark of proprietary interests is their endurance: a person who possesses a property right should not be asked to diminish their right to enjoyment of land, particularly when the community interests for which their rights must be relaxed are vague and ill-defined. A shift to damages would alter the way in which the principle of reasonable user, characterised by ‘give and take’ between neighbours, is conceptualised. Injunctions generally

⁷⁰ *ibid* [180].

⁷¹ Paul Davies, ‘Injunctions in Tort and Contract’ in Sarah Worthington and Graham Virgo (eds), *Commercial Remedies: Resolving Controversies* (CUP 2017) 128.

⁷² *Shelfer* (n 66) 322.

⁷³ Martin Dixon, ‘The Sound of Silence’ [2014] Conv 79, 84.

return a claimant to their position pre-nuisance. If damages are deemed the appropriate remedy, the courts have established a conceptual paradox, one in which the defendant's actions may be found to constitute a nuisance (the defendant's behaviour is unlawful), but the balance of 'give and take' has nevertheless been permanently altered in their favour, effectively giving the defendant stronger rights than their neighbour.⁷⁴

Even if public interests were defined, ranging in *Coventry* from 'the fact that a number of the defendant's employees would lose their livelihood',⁷⁵ to public use or enjoyment of the stadium,⁷⁶ planning permission does not weigh the strength of private rights against substantive public interests. A grant of planning permission 'should not be regarded as a manifestation of the wish of any authority'.⁷⁷ The permission is *negative* in nature in that 'the authority is clearly not *endorsing* the developments, but merely indicating that it *does not object to them*'⁷⁸. There is no careful and considered merits-based balancing exercise undertaken as part of the planning permission process. As such, there is no guarantee that such interests are as adequately balanced against private entitlements within administrative decision-making, as Lord Neuberger suggests.⁷⁹ Third parties that are likely to be affected by the planning permission, although entitled to make

⁷⁴ Emma Lees, 'Lawrence v Fen Tigers: Where Now For Nuisance?' [2014] Conv 449.

⁷⁵ *Coventry* (n 10) [124].

⁷⁶ *ibid* [239].

⁷⁷ JE Penner, 'Nuisance and the Character of the Neighbourhood' (1993) 5 *Journal of Environmental Law* 1, 24.

⁷⁸ Donal Nolan, 'Nuisance, Planning and Regulation: The Limits of Statutory Authority' in Andrew Dyson, James Goudkamp and Frederick Wilmot-Smith (eds), *Defences in Tort* (Hart Publishing 2015) 194 (emphasis in original).

⁷⁹ *Coventry* (n 10) [91].

representations to the planning authority about a proposed development, have constricted standing in the process. Any challenge to a planning authority's decision must be brought by way of judicial review,⁸⁰ which in any case does not lead to a consideration of merits. Moreover, some public interests identified in the case law—such as the development of the Docklands in *Hunter v Canary Wharf*⁸¹ or the provision of sewerage infrastructure in *Marcic v Thames Utility Council*⁸²—are not always severable from private interests and are in fact sometimes intrinsically linked to private profit-making.⁸³ Allowing private law property rights to be relaxed by relying on the assumption that there is a guarantee of public interest protection, or indeed that any public interest can be easily isolated from competing private interests, is therefore misplaced. In the same vein, the administrative nature of decision-making should not, in any case, be a source of the curtailment of private law rights. Although, again, this is not a curtailment *per se* of an individual's rights, as planning permission is not relevant to the question of liability, it has been established that the content of an individual's proprietary interests is nevertheless diluted by the relaxation of *Shelfer*. Watering down private law rights in this way should only be done with an expression of *legislative*, as opposed to *administrative*, will in the form of statutory authority. As noted by Peter Gibson LJ in *Wheeler v Saunders*, the courts 'should be slow to acquiesce in the extinction of private rights without compensation as a result of administrative decisions which cannot be appealed and are difficult to challenge'.⁸⁴

⁸⁰ Town and Country Planning Act 1990 (UK), s 78.

⁸¹ [1997] UKHL 14.

⁸² [2003] UKHL 66.

⁸³ Lee (n 66) 986.

⁸⁴ [1996] Ch 19 (CA) 28, 35.

The increased role of planning permission in the determination of remedies has demonstrated a clear movement towards the prioritisation of the public interest. The new orthodoxy signalled by *Coventry* not only undermines the essence of the claimant's right to enjoyment of land, but through its liberalisation on the grant of damages, also threatens to re-appraise the nature of the tort of nuisance as a tort against *land* altogether. Regrettably, the Supreme Court was overly cautious, declining to fully clarify how cases involving public interest and planning permission may be decided in future, while at other times too readily embracing a new paradigm by reference to the public interest, without having first established a robust framework to ensure certainty, opting instead for a case-by-case approach. In order to restore balance, the law of private nuisance must return to its emphasis on the endurance of property interests, as opposed to relying on planning permission to determine the strength with which these interests can be asserted. A potential framework is suggested below to highlight how the courts may approach this issue in the future, accounting for the concerns that have been discussed above while ensuring that the integrity of the law of nuisance is maintained.

C. Scrutinising the Public Interest

The courts' reluctance to challenge the *substance* of administrative decision-making is understandable given their lack of expertise to *substantively* evaluate competing public interests. As Lee highlights, 'most would agree that a public sewerage system is a good thing, for example, but precisely how effective it needs to be, who should pay, and how much, raises equally plausible claims and high economic stakes'.⁸⁵ It is unlikely that the institutional

⁸⁵ Maria Lee, 'The Public Interest in Private Nuisance: Collectives and Communities in Tort' (2015) 74(2) CLJ 329, 347.

capacity of the courts extends to the determination of such issues. However, to concede to the purported public interest protection within planning permission without adequate scrutiny is to adopt an overly deferential attitude toward planning authorities, which in turn overlooks the courts' competence to scrutinise the *process* of such decision-making. It should not be forgotten that courts can and should employ their constitutional capacity to engage in a meaningful balancing exercise that gives due weight to private property rights. This exercise is not one which weighs the *substance* of these rights, but instead determines whether, on balance, the public interests professed to be taken into account were arrived at by the planning authority through a reasoned and justified *process*. Judges may 'consider the expertise and information available to and used by the decision-maker, and what the decision itself purported to do'⁸⁶, or 'enquire whether a cost-benefit analysis (or in other cases, a general weighing of pros and cons) was in fact carried out, and whether the decision-maker engaged reasonably with the interests of most relevance to the litigation'.⁸⁷ This avoids undue interference with administrative decisions, while placing a check against the uncritical adoption of planning permission into the determination of the remedies to be awarded in nuisance. In light of the certainty and endurance of private property rights, the decision to grant damages in lieu of an injunction should depend on the clarity of the public interests sought to be incorporated and the justifiability of the confidence sought to be placed in the planning authority for reaching such conclusions.

⁸⁶ *ibid* 352.

⁸⁷ *ibid* 353.

Conclusion

The preceding discussion has revealed that tort law continues to carve out more space for the public interest at the expense of individual entitlement; the law continues to grapple with definitional problems because of the effort spent addressing the private and public domains as if they are immiscible interests. They need not be in diametrical opposition to one another. Even when the interests do conflict, the private interest must remain the starting point. For defamation this is the reputation interest, whereas in nuisance this is the property right. From there, the public interests at play within the specific facts of a case must be clearly *identified, enumerated, and defined*. The key here is to increase the transparency (ensuring comprehensiveness in identification) and precision (ensuring clarity in definition) with which the courts engage in their analysis. It is only from this vantage point that they can appropriately balance the relevant interests in a case, whether they are competing or complementing one another, or whether they overlap or are mutually exclusive.

This contextual approach stands in contrast to other methods, previously mentioned, that begin with the prioritisation of one *type* of interest over another: the social aims approach prioritises public interests while the essentialist approach *always* prioritises private interests. The approach suggested in this article is preferable because takes a rounded and nuanced account of the interests involved within a dispute, taking the private interest as the starting point, but being appropriate deferential when a significant public interest is at play. Reaffirming the importance of private rights means that any encroachment upon them ought to be fully justified by reference to precisely defined public interests. This will bring some much-needed clarity to the nature of the competing interests in all tortious disputes.

Overcoming the Law's Marriage to the Past

A Re-Assessment of the Meaning of 'Reasonable Steps' in Non-Commercial Guarantees following Royal Bank of Scotland Plc v Etridge (No.2)

Merit Flügler*

Abstract—This article considers the law surrounding non-commercial guarantees and the steps that must be met in order to ensure that a third party's wrongdoing does not affect the validity of the guarantor's contract. Non-commercial guarantees often arise in mortgage and security relationships entered into by married couples, where the husband unduly influences the wife to sign a guarantee for his debt. It is suggested here that in order for the law to realign itself with contemporary societal standards, more stringent requirements need to be imposed on lenders and solicitors. It will be argued that this includes (a) private meetings with the guarantor/wife, (b) independent legal advice from

* Merton College, Oxford. I am grateful to Mindy Chen-Wishart for her continuous support and the OUULJ editorial board for their comments on earlier drafts. Any deficiencies and errors remain my own.

different solicitors, and (c) a lower threshold for what is required in order for a solicitor to veto a guarantor transaction.

Introduction

In cases of non-commercial guarantees, where the relationship between the guarantor and debtor is personal and thus open to influence, there is grave scope for abuse. *Royal Bank of Scotland Plc v Etridge (No.2)*¹ (*Etridge*), while taking a substantial step in the right direction—holding that a guarantee is unenforceable if the lender has not taken reasonable steps to ensure that the guarantor was properly advised—still has a long way to go; we must ask ourselves: is *Etridge* the most desirable solution to resolving the three-party problems of relationship influence? The optimal solution might very well be contingent on contemporary morality. As Lord Bingham noted in *Etridge*, equity ‘is not past the age of child-bearing’.² In His Lordship’s opinion, ‘the existence of this obligation in all non-commercial cases does not go beyond the reasonable requirements of the *present times*’.³ This article will argue that as twenty years have passed since *Etridge*, the reasonable requirements of equity have changed in this time. The law has gradually become more protective of the interests of the weaker party—a development which has been described as one of the cornerstones of the evolution of contract law⁴—and lenders have also begun to incorporate principles of socially responsible lending into their practises. Because of society’s shifting views, the law now falls short of equity’s ‘reasonable requirements’, and certain reforms are necessary.

This article will take the following approach. First, the societal and practical background surrounding non-commercial

¹ [2001] UKHL 44.

² *ibid* [89].

³ *ibid* (my emphasis).

⁴ Hondius, ‘The Protection of the Weak Part in a Harmonised European Contract Law: A Synthesis’ (2004) 27 *Journal of Consumer Policy* 245.

guarantees will be considered in order to contextualise the applicable law and propositions for reform. Next, *Etridge's* three-stage approach will be explained. Then, the following argument will be made: (1) There is currently a divergence between the law and societal norms. (2) The law ought to accord with contemporary morality and societal norms. (3) If the lender fails to adhere to these moral and social norms, they are not 'innocent'. (4) If the lender is not innocent, they do not merit the protection of the law and the contract should be voided. (5) In order to bridge the law-morality divide and thereby ensure that banks adhere to the moral standard, certain reforms are necessary, including more stringent steps for banks to follow prior to granting a loan, such as the provision of private meetings, as well as a requirement of independent legal advice. Lastly, and by analogy with unconscionable bargains, it will be argued that the solicitor's role in the process should also be reformed.

1. A Moving Picture

A. Context

The most common scenario where guarantor problems arise is in a relationship between husband and wife, as was the case in *Etridge*. In such a scenario, the husband needs a loan for personal or business purposes, and due to his own indebtedness or low credit score, the bank is unwilling to lend to the husband without some kind of guarantee on repayment. To this end, the wife is asked by her husband to serve as a guarantor and if successful, the bank will normally charge the land owned by the borrower.

This can include the guarantor entering a mortgage.⁵ Consequently, if the husband is later unable to repay the loan, it is the guarantor (the wife) and their family who are forced into debt and homelessness, because the bank will likely take possession and sell the property with vacant possession, forcing its occupants to leave the home.⁶ Other non-commercial guarantee relationships exist, such as agreements between employees⁷ and friends,⁸ but they are less common. The nature of the relationship matters because the husband-and-wife cases have the widest scope for abuse; they are the most obvious examples of relationships of influence: married couples will rely on each other for advice and try to accommodate each other's needs. However, the commonality that links all such relationships is that the guarantor's only reason for signing the contract and agreeing to act in such capacity is personal: by definition (a 'non-commercial' guarantee), the guarantor cannot have any involvement in the business activity of the debtor (the husband).

This husband-wife scenario is more common than one might think.⁹ In *Barclay's Bank plc v O'Brien*¹⁰ (*O'Brien*), Lord Browne-Wilkinson noted that there had been eleven reported cases in the Court of Appeal in the last eight years, and in *Etridge*

⁵ This was the case in *Etridge*, where the eight joint appeals concerned land mortgaged to lenders.

⁶ Power to sell is given under the Law of Property Act 1925, s 101(1)(i), exercisable after three months if the whole sum is due (s 103(i)), after two months if the interest is due (s 103(ii)), or if there is a breach of the mortgage agreement (s 103(iii)). The right to possession always exists following *Four Maids Ltd v Dudley Marshall* [1957] 1 Ch 317, 320.

⁷ *Credit Lyonnais Bank Nederland NV v Burch* [1997] 1 All ER 144, [185]-[186].

⁸ *Banco Exterior Internacional SA v Thomas* [1997] 1 WLR 221.

⁹ Jack Beatson, *Anson's Law of Contract* (31st edn, OUP 2016).

¹⁰ [1994] 1 AC 180.

it was noted that the House of Lords had heard appeals in eight separate cases on the topic.¹¹

B. The Current Law

The current law was set out in *Etridge*, restating the *O'Brien*¹² principle (the bank's constructive notice of wrongdoing) and extending it to all guarantees entered into by guarantors and principal debtors who do not have a commercial relationship.

Etridge held that 'reasonable steps' are required of the lender if they desire to enforce a guarantee affected by vitiating factors. A guarantee is unenforceable if: (a) there is a vitiating factor (undue influence, misrepresentation etc.) which affects the dealing and communication between the guarantor and the debtor, (b) the bank knows that the guarantor is acting in a non-commercial manner and that the transaction is for the benefit of the primary debtor, and (c) the bank has not taken 'reasonable steps' to ensure that the guarantor was properly advised. In detail, this means:

- (a) The first 'stage' necessary for a lender to enforce a guarantee is the absence of vitiating factors against the primary debtor—if no vitiating factor or 'misconduct of a third party'¹³ exists, the bank can *always* enforce the guarantee. However, the existence of any such factors between the debtor and guarantor lays the foundation for the unenforceability of the contract between guarantor and

¹¹ *ibid* 185.

¹² For discussion of the principle, see Mindy Chen-Wishart, 'The O'Brien Principle and Substantive Unfairness' (1997) 56 *CLJ* 60.

¹³ *Etridge* (n 2) [30].

debtor. Vitiating factors include undue influence as well as misrepresentation.¹⁴

- (b) The next stage is the lender's notice of the vitiating factors. The law had already undergone development before *Etridge*. The initial position required the lender's *knowledge* of the vitiating factor(s) in order to render the guarantee unenforceable.¹⁵ The *O'Brien* principle then extended this doctrine to include constructive knowledge, or as Lord Browne-Wilkinson put it, the lender's 'constructive notice' of the debtor's wrongdoing.¹⁶ Constructive knowledge was said to be present where there was (i) a transaction with no financial advantage to the wife, and (ii) 'a substantial risk in transactions of that kind that the husband had committed legal or equitable wrongs, such as unduly influencing his wife.'¹⁷ Lord Browne-Wilkinson's constructive notice test has been stripped from the law by *Etridge*, on the basis that the law must avoid arbitrary distinctions between, for example, people in sexual relationships and those in platonic relationships,¹⁸ and also must avoid banks being forced to investigate the kind of relationship that existed between debtor and guarantor.¹⁹ The *O'Brien* principle was replaced with the simple test that the bank must take 'reasonable

¹⁴ *Annulment Funding Co Ltd v Covey* [2010] EWCA Civ 711 [64].

¹⁵ *O'Brien* (n 10).

¹⁶ *ibid* 195-196. This use of the term was itself found to be an extension, see *Etridge* (n 2) [42]: 'Lord Browne-Wilkinson acknowledged he might be extending the law'.

¹⁷ *O'Brien* (n 10) 196.

¹⁸ *Etridge* (n 2) [86].

¹⁹ *ibid* [87].

steps' in every case where the relationship between the guarantor and the debtor is non-commercial.²⁰

- (c) The final stage is the completion of 'reasonable steps' in order to make the agreement enforceable.
- (i) The burden on the banks is relatively simple, requiring the guarantor to seek legal advice. The bank must then seek written confirmation from the guarantor's solicitor that they have explained the nature and effect of the lending documents to the wife.²¹ The court expressly rejected requirements of private meetings between the bank and the guarantor,²² requiring the bank to attempt to discover undue influence itself or obtaining confirmation that the wife was not unduly influenced²³ into signing the contract (in contrast to confirmation that the wife was instructed). These rules are displaced if the bank knows that the guarantor was not duly advised or 'if the bank knows facts from which it ought to have realised that the wife has not received the appropriate advice'.²⁴
- (ii) Independent advice for the guarantor is not necessary.²⁵ It is for the guarantor to choose whether they want to be represented by the same solicitor as the debtor, and given that solicitors are often shared in the common husband-wife cases, it is unlikely that the guarantor will go out of their way to find a new solicitor.

²⁰ *ibid* [44], [87].

²¹ *ibid* [51].

²² *Etridge* (n 2) [55], [94].

²³ *ibid* [53].

²⁴ *ibid* [57].

²⁵ *ibid*.

As an addendum, in situations where the creditor knew or had reasons to believe that the guarantor already knew of the risks associated with the situation, there is no obligation to follow these steps to begin with.²⁶

C. Law and Societal Norms

When thinking about the divergence of law on the one hand, and societal standards and norms on the other, we are reminded of Lord Bingham's reference to equity not being past the age of child-bearing and its contingency upon current societal norms. The importance of responsible lending is increasingly being recognised in legislation. In 2014, the Financial Conduct Authority introduced the Consumer Credit sourcebook (CONC).²⁷ The sourcebook includes provisions such as CONC 5.2A.31 and CONC 5.2A.32, which discuss the factors to be taken into account in the debtor's creditworthiness assessment by the lender where there is a guarantor. CONC 5.2A.32 (5) says that the lender must have regard to information such as the total potential liability of the guarantor under the guarantee and any other potential adverse consequences for the guarantor arising under the guarantee from a failure to make a payment by the due date, when making the assessment. Secondly, and again in 2014, the European Union introduced the EU Market Directive 2014/17/EU, laying down a more regulated framework for the sphere of consumer mortgage credit agreements and residential immovable properties. These provisions are built on the foundations of responsible lending principles. They aim to prevent the bankruptcy of the borrower by ensuring that the

²⁶ *ibid.*

²⁷ Karen Fairweather, 'The development of responsible lending in the UK consumer credit regime' in James Devenney & Mel Kenny (eds), *Consumer Credit, Debt and Investment in Europe* (CUP 2012) 84.

lender's terms are fully understood, that the loan is affordable, and that alternative arrangements can be made for unforeseen circumstances. This type of legislation is clear evidence of a shift in lending culture.

Moreover, the importance of promoting socially responsible lending has been recognised by the banks themselves. So far, lenders have taken a more substantive and less procedural approach to incorporating responsibility principles, paying more attention to environmental, social and governance (ESG) criteria that are traditionally used by investors to examine and evaluate potential investments.²⁸ The importance of ESG factors can be seen in a 2019 survey of 194 credit risk experts working in banking and finance: 86% of the participants believed that ESG factors were becoming more important thanks to increased investor demand, and 83% found that ESG factors were playing an increasingly crucial role in credit risk areas.²⁹ It is thus evident that the banking sector is becoming increasingly aware of their social responsibility to their customers.

There has also been international recognition of the importance of corporations examining their internal practices and following recognised guidelines for social governance, and although not binding on UK lenders, lessons can nevertheless be learnt. The 2011 Organisation for Economic Co-operation and Development (OECD) guidelines for responsible business

²⁸ There are some differences between equity capital and corporate lending. Ralph De Haas and Alexander Popov, 'Finance and carbon emissions' (2019) ECB Working Paper Series 2318 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3459987> accessed 21 January 2021.

²⁹ S&P Global Market Intelligence, '2019 ESG Survey', <https://pages.marketintelligence.spglobal.com/ESG-Survey.html> accessed 15 February 2021.

conduct (the Guidelines) ask businesses to (i) make a positive contribution to economic, environmental, and social progress with a view to achieving sustainable development; and to (ii) avoid and address adverse impacts through their own activities and seek to prevent or mitigate adverse impacts directly linked to their operations, products, or services by a business relationship.³⁰ Both principles can be applied to banks, and more specifically, to non-commercial guarantees. Interestingly, if a lender were to fail to prevent undue influence in a non-commercial guarantee, it would not meet part (ii) of the Guidelines. The Guidelines therefore provide a strong contextual framework for what we can ask of lenders and what their duties are in order to be ‘innocent’ in the transaction. This has in some forms already been recognised by individual banks, such as Lloyds Bank’s introduction of a Domestic and Economic Abuse Team.³¹

Observing these changes in legislation, ESG criteria and international guidelines together, there is an undeniable trend towards more socially conscious lending practices, aimed at protecting the weaker party even when this might come at a cost or inconvenience to the bank. Such an evolution in principles should be reflected by a higher threshold for the ‘reasonable steps’ that lenders must follow.

³⁰ OECD (2011), ‘OECD Guidelines for Multinational Enterprises’, <<http://www.oecd.org/daf/inv/mne/48004323.pdf>> accessed 29 December 2020.

³¹ Lloyds Bank, ‘Lloyds Bank 2020 Annual Report’ <<https://www.lloydsbankinggroup.com/assets/pdfs/investors/annual-report/2020/2020-lbg-annual-report.pdf>> accessed 29 December 2020.

D. 'Innocent' Lenders

The main argument as to why the guarantee contract ought not to be undermined is because the lender is 'innocent' in the transaction, as the lender does not participate in the wrongful undue influence that renders the contract voidable. It is a well-established rule in contract law that if a third party in good faith acquires an interest in property, a merely voidable transaction can no longer be voided.³² This also applies to the rescission of contracts for undue influence. It is, however, questionable to what extent one can really say that law in relation to non-commercial guarantees requires the bank to act in 'good faith'. Good faith requires a party to be open, fair, and honest. According to *Etridge*, banks can comply with what is required of them by law (i.e. comply with the 'reasonable steps'), yet still fall far short of the moral standard for good faith. There might even be instances where calling the lender reckless, indifferent, or negligent is a more accurate description of their conduct.

Thus, in order to be a *fully 'innocent'* lender, the bank must incur no *moral* culpability by meeting their *moral* duties and responsibilities. The question of whether the moral duty is met should be the same as whether the legal duty is met, and if the answer is negative, the contract should be nullified.

The preferred reasoning is already set out in *O'Brien*: the issue is not the advantage-taking by the bank, but the notice.³³ A lender cannot be acting responsibly or innocently when ignoring the abuse of the contractual partner. Thus, it is contrary to the principles of contract law to maintain a contract³⁴ where the third party has notice of the undue influence. In *O'Brien*, Chen-Wishart

³² *Phillips v Brooks Ltd* [1919] 2 KB 243.

³³ *O'Brien* (n 9) 195-196.

³⁴ *ibid.*

characterises the question to be asked as being ‘whether the conduct of the creditor has fallen short in any respect which would amount to *unconscionable disregard* for the predictable vulnerability to the misfeasance or influence of the debtor’.³⁵ The ‘language of the jurisdiction’ is one of notice.³⁶

2. ‘Reasonable Steps’?

A. *The Role of the Bank*

The current requirements fail to acknowledge that in order to merit protection, the lender needs to be *fully innocent*. Several new steps and safeguards should be introduced to ensure that the law governing non-commercial guarantees is in line with principles of contractual fairness and, as seen above, heightened contemporary standards of lender responsibility.

I. *A Plea for Private Meetings*

Private meetings between the lender and the guarantor ought to be included as a ‘reasonable step’, rather than the joint meetings currently required. The procedure of the meetings should remain the same, but the guarantor and debtors should have to individually meet with employees of the bank. This suggestion was discussed and rejected in *Etridge* on the basis that the cost of private meetings precluded them from being made mandatory. On closer inspection, this objection does not hold water.

³⁵ Mindy Chen-Wishart, ‘The O’Brien Principle and Substantive Unfairness’ (1997) 56 CLJ 60 (my emphasis).

³⁶ *ibid.*

On one hand, costs ought to be kept to a minimum; otherwise, the lender will pass on the costs to borrowers and the availability of credit will suffer. This would harm all borrowers as a group, demonstrating why it is wrong to portray such costs as a conflict merely between the interests of the lenders and of the wronged guarantor.³⁷ Opponents to the private meetings proposal might therefore argue that it is an expensive burden to train the lender's staff to facilitate these meetings. Staff costs make up a significant portion of the operating expenses of banks—for example, such costs made up 26% of expenditure at Lloyds Bank in 2020.³⁸ However, and at the end of the day, the suggested reform simply means that a meeting which would be conducted anyway will now be held twice: once for the guarantor, once for the debtor. The training that is currently offered entails, on average, 40 hours at 1200 pounds per head per year.³⁹ One extra hour of training for the staff dealing with the guarantor (costing only around an extra 30 pounds per head) could make a significant difference, in that they would then be likely to perform any additional tasks linked to the guarantor contract with greater care. Moreover, some firms may already have similar training in these areas—such as Lloyds Bank's domestic abuse awareness training—which would allow the training programme to be expanded upon at minimal extra cost.

Next, two evaluations can be made about the potential costs of litigation. First, as above, qualified staff can prevent or at least mitigate this issue. Second, the problem of litigation is present both on the side of the bank and on the side of the

³⁷ See Antony Dnes, *Principle of Law and Economics* (3rd edn, Edward Elgar) 121.

³⁸ Lloyds Bank 2020 Annual Report (n 31).

³⁹ Training Magazine, '2019 Training Industry Report' <https://trainingmag.com/sites/default/files/2019_industry_report.pdf> accessed 06 January 2021.

claimants. If the guarantor is unduly influenced and ends up homeless, resultant litigation is likely. As the law is forced to choose between two innocent parties, it is only right that the bank should carry this burden of cost, for two reasons. Firstly, there is a significant difference in resources: banks usually have greater access to lawyers and funds than a non-commercial guarantor. Secondly, there is the difference in bargaining power: bank loans will often be ‘take it or leave it’ contracts; it is a business-consumer contract with a large power imbalance in favour of the bank. As McBride neatly summarises it, the bank is better equipped to take on the risk of litigation.⁴⁰ Even when choosing between two innocent parties, there are still different degrees of ‘innocence’. The guarantor, by default, is completely innocent. However, the lender’s innocence, as examined above, can vary. When deciding between a resourceful stronger party with some degree of non-innocence and a weaker, financially vulnerable counterparty, it is contrary to both morality and doctrine if the latter is not protected.⁴¹ Thus, the chance of litigation, although a recognised cost, is justifiably shifted onto the lender.

It was also stated in *Etridge* that private meetings would be a ‘disproportionate response to the need to protect those cases,

⁴⁰ Nicholas McBride, ‘Random Lines – Land Law’ (*McBride’s Guides*, 13 August 2013) <<https://mcbridesguides.com/category/random-lines/land-law/>> accessed 28 April 2021.

⁴¹ Evident in cases surrounding unconscionable bargains such as *Fry v Lane* [1880] 40 Ch D 312, 322, where Kay J articulated that equity most commonly intervenes in cases of a party ‘just of age, his youth being treated as an important circumstance’, or where the vendor is ‘a poor man with imperfect education’. These principles are also visible in *Cresswell v Potter* [1978] 1 WLR 255. For wider academic discussion, see Waddams, ‘Protection of Weaker Parties in English Law’ in Mel Kenny, James Devenney & Lorna Fox O’Mahony (eds), *Unconscionability in European Private Financial Transactions: Protecting the Vulnerable* (Cambridge University Press 2010).

presumably a small minority, where a wife is being wronged.⁴² This shows a flawed prioritisation of the interests involved. This calculation of proportionality does not seem to take into account the severity of the resultant harm: even though the harm is not likely, its effects on the wronged individual are immense. As explained above, homelessness and insolvency of the guarantor are common results in cases where the debtor defaults. Given the obvious need to prevent such an outcome, calling a small increase in costs a ‘disproportionate response’ seems almost cynical.

There may be non-pecuniary feminist challenges to the argument for private meetings which may characterise greater and stronger intervention as a patriarchal assumption of women’s inability to self-protect. However, this is not a question of patronising women and assuming they are unable to make wise decisions relating to marriage and their financial affairs; it is merely a recognition of the fact that there is scope for abuse. There are three further reasons why such a change would not be anti-feminist. Firstly, for women who are resilient, privileged, and capable enough to resist any influence, they suffer no harm, and the reforms do not present any great additional burden. Secondly, the reforms do not only apply to women: *Etridge* made clear that the ‘reasonable steps’ affect both heterosexual and homosexual relations, cohabitational or not.⁴³ Although the issue might be more prevalent in husband-wife situations, the need to rebalance the priority given to the interests of the banks over the interests of the weaker party is not gendered. Thirdly, for those women who are more likely to be influenced, greater protection can only be a benefit. Thus, the argument that such legislation is

⁴² *Etridge* (n 2) [61].

⁴³ *ibid* [86].

antifeminist is flawed to the extent that the inverse seems more likely.

Evidently, the objections to private meetings fall down on closer inspection. We can now also consider the benefits of private meetings, further demonstrating their necessity. Firstly, private meetings would make the existing protective mechanisms more effective. If insufficient attention is paid to the debtor and guarantor by the bank in the meeting, the result is that the meetings will be treated as merely performative, thereby undermining the effectiveness of the other procedural safeguards. Later steps, such as the provision of independent advice, are less successful, as the wife may well not understand why that later step is desirable or necessary. Thus, by rejecting the status quo and introducing private meetings, there is a lower likelihood that insufficient attention will be paid to both the debtor and guarantor by the bank, decreasing the chance of performative meetings. Private meetings would also provide more room for staff to identify misunderstandings. The current law is that if it is clear that the guarantor is under a false impression as to the terms of the agreement due to misrepresentation or another vitiating factor, it is on the lender to clarify that misunderstanding. Private meetings give the lender the opportunity to rectify any issues before the process goes awry. Again, private meetings should be seen as laying the foundations for the success of subsequent safeguards. Private meetings would allow the guarantor to ask questions more freely. Without the imminent fear and intimidation caused by the presence of the potentially influential or abusive debtor, the guarantor is in a better place to make enquiries about the terms, express discomfort, or seek to renegotiate the contract. In most cases, this will not be possible if the partner that is exerting pressure on the guarantor is present in the room. Separation will also eliminate potential conflicts of interest in the eyes of the (potentially) abused party. This way, the

guarantor will know that it is their interest that is being guarded, and that the bank's employee is there for their benefit, rather than the benefit of their partner.

Here, an analogy can be drawn to medical practice. When a person seeks advice on the termination of pregnancies, such as by an abortion pill, the accompanying party is asked to wait outside so that the affected party can be advised separately.⁴⁴ The reason for this is to avoid reproductive coercion.⁴⁵ Although this example takes place in a different field, it should be acknowledged that separate and private consultation is already common practice in other areas, in order to avoid the same kind of manipulative pressure that a debtor may place on the guarantor.

It is therefore pressingly clear that the private meeting objections raised in *Etridge* are not watertight. When we also consider the potential benefits that such meetings could bring, it is surprising that the House of Lords did not consider the case for private meetings more closely. Their introduction, far from being a 'disproportionate response', would only act to catch potential victims of abuse slipping through the procedural net.

Finally, we must also note that the current standards for the procedure and content of meetings need to be enforced *in practice*. Time and time again there has been criticism that meetings are seen as purely performative, rather than being taken seriously and complying with all the standards required.⁴⁶ This includes communicating in comprehensible language, double checking

⁴⁴ British Medical Association, 'The law and ethics of abortion – BMA views' <<https://www.bma.org.uk/media/3307/bma-view-on-the-law-and-ethics-of-abortion-sept-2020.pdf>> accessed 1 May 2021.

⁴⁵ Grace & Anderson, 'Reproductive Coercion: A Systematic Review' (2018) 19(4) *Trauma Violence Abuse*, 371.

⁴⁶ *Etridge* (n 2) [33].

whether the contracting parties truly understand the nature of the terms etc. If the standards continue to be only applied in a manner that reduces them to mere formalities, there may need to be further reforms, such as the introduction of an independent accountability mechanism, in order to force compliance.

B. Independent and Separate Legal Advice

As well as the need for private meetings between lender and guarantor, the advice the guarantor receives from their solicitor ought to be *independent*, and therefore *private* too. Currently, all the bank needs to do is obtain confirmation from the guarantor's solicitor that the guarantor has been advised about the nature and effect of the transaction,⁴⁷ and all the debtor and guarantor need to do to meet their obligation is to attend a meeting with their solicitor. The solicitor may act for the guarantor's husband, or even the lender itself, leaving the contractual door wide open for relationship abuse. Whereas under the current law the legal advice can be given in the presence of the spouse,⁴⁸ reform would allow for distance to the abuser and prevent further influence and intimidation during the meeting.

Similar to the separation argument for private meetings, meeting the solicitor without the debtor is advantageous in that the spouse has greater freedom (and will feel they have this greater freedom) to honestly discuss whether they want to enter the transaction at all, or on what terms. Receiving legal advice from different people will also ensure that the guidance given is independent and free from any conflicts of interest—any potential conflict is not just *reduced*, but *eliminated*. As former Lord Chief Justice Hewart famously exclaimed, 'justice should not only be done, but should manifestly and undoubtedly be seen to be

⁴⁷ *Etridge* (n 2).

⁴⁸ *ibid* [66].

done'.⁴⁹ In the context of advising abuse victims, this is particularly true; if it is not *abundantly clear* to the victim that there is confidentiality and separation, they will not be able to benefit from it—having professional standards of conduct that the victim knows nothing about is of no use. Justice must be seen to be done. This recommended change mirrors the reasons for private meetings with the lender's staff; the two safeguards complement each other and jointly provide for a more effective safety net.

Admittedly, requiring all guarantors to find separate counsel from their spouse will require greater effort on the guarantor's part. In most cases, married couples will have the same solicitor that they turn to for legal advice. Thus, it is an additional burden for them to find different solicitors. However—and in contrast to the earlier argument about costs being justifiably shifted onto the lender—in order to create a level playing field, it is only reasonable to put some additional burden of undue influence prevention onto the debtors and guarantors too. Moreover, hearing a second independent opinion will never actively disadvantage the guarantor. Lastly, such an inconvenience can be mitigated: the family solicitor is likely to be a member of a firm of solicitors, and it is not particularly troublesome to go to a different solicitor (in the same building).

Moreover, as Fuller argues, formalities have a cautionary function.⁵⁰ Finding separate legal counsel gives the guarantor more time to reassess the situation and discuss their concerns with their confidantes. This additional consideration time helps prevent them from signing something they ought not to. This method will help all guarantors, including those considering

⁴⁹ *Rex v Sussex Justices* [1924] 1 KB 256 [259].

⁵⁰ Fuller, 'Consideration and Form' (1941) 41 *Columbia Law Review* 799.

signing something for less serious, but equally irrational reasons, such as mere lust or temporary endearment.

Nevertheless, though independent legal advice and private meetings with those advisers are necessary, they are not always sufficient. In relationships of pressure, fear and intimidation, the abusive and dominant partner will not need to be present in order to influence the wife's behaviour. Other steps can and should be taken to counteract the effect of undue influence. The next step that should be taken is raising the threshold for what counts as 'adequate advice', discussed below.

I. The Problematic Role of the Solicitor: Adequate Advice

One of the biggest issues the wronged guarantor will face is the fact that the lender can escape liability by passing on all responsibility to the solicitor.⁵¹ If the solicitor's advice is inadequate, there is the possibility of a tort action for professional negligence against the solicitor, but there is no effect on the contract between the lender and the guarantor. A tort claim against the solicitor is not an effective alternative to voiding the contract: as McBride points out, because the bank is able to enforce the contract, the last thing the guarantor is able to do is sue,⁵² as she ends up both homeless and penniless. The only way the solicitor can currently protect the guarantor from enforcement of the guarantee is by declining to act for the bank any further. The solicitor has access to a de facto 'veto', but it can only be used in cases 'where it is glaringly obvious that the wife is

⁵¹ Morris, 'Wives are Told: Don't Blame the Bank, Sue Your Solicitor: Royal Bank of Scotland v Etridge (No. 2)' (1999) 7 Feminist Legal Studies 193.

⁵² McBride (n 40).

being grievously wronged'.⁵³ This threshold is too high. It should be reduced in order to offer a more holistic protection of the guarantor.

II. Reducing the Threshold of the Solicitor's 'Veto'

There should be a lower threshold to allow a solicitor veto, as the bar is currently set so high that it effectively absolves the solicitor of any responsibility in preventing undue influence. Firstly, the wrong suffered must be 'grievous', indicating that the degree of undue influence present must be *really bad*. Secondly, the influence must be 'glaringly obvious', meaning that the undue influence not only needs to be glaringly obvious, but it must *also* be '*glaringly obvious*' that the undue influence is '*grievous*'. Unless the abuse quite literally jumps into the face of the solicitor, there is no further protection at this step.

Moreover, comparison with the undue influence case law suggests the threshold for the solicitor veto should be reduced. Particular reference should be made to *Crédit Lyonnais v Burch*⁵⁴ (*Crédit Lyonnais*), which embodies the principle that advice is inadequate if the guarantor enters into a transaction with no sensible solicitor could have recommended. In the context of non-commercial guarantees, no inference about the quality of advice can be drawn by the court from the outcome, no matter how grossly unfair the guarantor's ultimate position. This does not follow the line taken in undue influence cases. The relationship between the quality of advice and the outcome is the same in both scenarios, but the law concerning each area is divergent. There appears to be no reason in principle for such an inconsistent distinction and the law, in cases of non-commercial

⁵³ *Etridge* (n 2) [61]-[62].

⁵⁴ [1997] 1 All ER 144.

guarantees, ought to recognise this by following the protection given in undue influence cases.

C. A Comparison to Unconscionable Bargains

The law of unconscionable bargains has a similar doctrinal structure to non-commercial guarantees. An unconscionable bargain also defeats a contract based on the unfairness of the transaction. Following *Fry v Lane*,⁵⁵ three elements are necessary to cause contractual vitiation:⁵⁶ (a) the weaker party must be at a serious disadvantage, i.e. they must have a bargaining impairment; (b) the stronger party must exploit the disadvantage in a morally culpable manner; and, (c) the contract that is entered into must be oppressive, meaning the transaction ‘must shock the conscience of the court’, in that there is a *substantial* undervaluation of the consideration offered by the weaker party.⁵⁷ The similarities between these vitiating factors and those in non-commercial guarantees are striking: in both cases, the ability of the weaker party to negotiate or strike a good bargain is affected. In unconscionable bargains, this is often due to an intrinsic bargaining impairment,⁵⁸ while in non-commercial guarantees the bargaining difficulties are induced through undue influence or misrepresentation by the debtor. Another similarity is the immoral conduct of the stronger party, which in non-commercial guarantees includes the debtor and the lender. Lastly, the resulting

⁵⁵ (1888) 40 Ch D 312.

⁵⁶ Beatson (n 9).

⁵⁷ *Alec Lobb (Garages) Ltd v Total Oil (Great Britain) Ltd* [1983] 1 WLR 87, 95.

⁵⁸ See *Creswell v Potter* [1978] 1 WLR 255, ‘less highly educated.’ It is however possible that it can also be external, as seen in *Backhouse v Backhouse* [1978] 1 WLR 243, 251, where a marriage break-up caused such emotional strain that an otherwise intelligent and competent person was impaired in their ability to bargain.

transaction will cause grave detriment to the weaker party in both cases. Although, unlike the law on unconscionable bargains, the resultant harm is not examined by the courts in non-commercial guarantee cases, in practice this discrepancy does not make a difference, as in both cases the weaker party is left with nothing without legal intervention.

Despite these similarities, the judicial treatments of the two situations differ. In unconscionable bargains, the role of adequate advice is more prominent and the requirements more stringent. Following *Bank of Montreal v Stuart*,⁵⁹ if the defendant recommended independent advice, but the complainant refused to obtain it, then such an action is interpreted as evidence of the seriousness of their bargaining impairment. No such requirement or 'safety net' exists within non-commercial guarantees. It has already been argued that independent legal advice should be required in all cases of non-commercial guarantees, and although one cannot expect that, every time the wife refuses to employ her own solicitor, it must be because she is being pressured and manipulated, downright refusal to comply is out of the ordinary, given that it is only a slight inconvenience, indicating that there might be ulterior pressures at play. If both the debtor and the guarantor had the best interest of the other party in mind and were acting in good faith, they would want the best possible counsel for their partner.

Moreover, in the context of unconscionable bargains, if the complainant has received advice, the court may still find that, because of the harshness of the transaction or for other reasons, the advice was inadequate.⁶⁰ Alternatively, the court may find that, because of the severity of the complainant's impairment, they

⁵⁹ *Bank of Montreal v Stuart* [1910] UKPC 53, [1911] AC 120.

⁶⁰ *Fry v Lane* (n 55).

were unable to benefit from the advice.⁶¹ Again, no such safety net exists within non-commercial guarantees, other than *Crédit Lyonnais*'s principle that a transaction might be so unbalanced that the willingness of the guarantor to proceed after independent legal advice is further evidence of undue influence.⁶² This check alone is insufficient as it will only be evidence for stage one of the guarantee process (the underlying vitiating factor), and is consequently unrelated to the reasonable steps that are required of the lender. Proving satisfaction of step one will often not be helpful to the guarantor, as an underlying vitiating factor alone cannot vitiate a guarantee. While helpful in two-party situations where the difficulty is proving vitiating factors, the principle will not assist in a three-party cases.

By examining *Crédit Lyonnais* again, but this time more closely, we can see the relationship between grossly unfair outcomes and independent advice. The Court held that in unconscionable bargain cases the bank should have *insisted*, rather than merely *recommended*, that the claimant obtain independent legal advice. This is directly comparable to non-commercial guarantees, but the onus on the bank is much higher. Moreover, Millet LJ recognised the limited effect that independent advice—on its own—will have, supporting the position that the independence of the advice delivered is a necessary part of a number of protective mechanisms. 'It is not sufficient that the solicitor has satisfied himself that the complainant understands the legal effect of the transaction and intends to enter into it. That may be a protection against mistake or misrepresentation; it is no protection against undue influence'.⁶³ This position is one that is far more stringent than the current non-commercial guarantees

⁶¹ *Boustany v Piggot* [1995] 69 P & CR 298.

⁶² McBride (n 40).

⁶³ *Crédit Lyonnais* (n 53) 156.

regime, further demonstrating the need for law reform in order to harmonise these different, yet highly similar, areas of contract law.

Conclusion

This article has shown that the current law surrounding non-commercial guarantees does not provide sufficient protection for guarantors and is in need of reform. The current three-part structure of the law—composed of underlying vitiating factors, the need for notice of the lender, as well as the completion of reasonable steps—is coherent and serves as a good foundation. However, the ‘reasonable steps’ the lender must take are not adequate. Both the moral obligations and societal expectations of the banks exceed what the law prescribes to them. More—and indeed stronger—safety nets are required in order to effectively prevent undue influence and misrepresentation. ‘Reasonable steps’ should include the requirement of private meetings with the lender for the guarantor, as well as independent advice from the solicitor. Finally, the role of the solicitor must be reconsidered by reducing the threshold of what is required in order for the solicitor to ‘veto’ the transaction. These changes will provide the protection that guarantors so desperately need and deserve, while also realigning the law with contemporary societal and legal standards.

*Saviour Siblings, Commercial
Organ Donation and Commercial
Surrogacy: Finding the Right
Balance between Acceptable
Instrumental Use and
Impermissible Commodification*

Bruno Ligas-Rucinski*

Abstract—This article analyses the UK legal framework on three related topics, with particular focus on the ethical issues they raise: (1) the donation of human tissue by children to be used in the treatment of sick siblings with matching tissue types (saviour siblings), (2) the commercialisation of organ donation and (3) the commercialisation of surrogacy. After setting out the relevant law on these issues, the article will consider the rationales behind protecting the body from commodification in the first place. It shall then apply these rationales to the specific debates

* Recent graduate from Christ Church, Oxford. This article was written whilst I was an undergraduate. I am grateful to the tutors at Christ Church, past and present, for their unwavering support during my studies. I am also grateful to the OUULJ editorial board for their comments on earlier drafts and Elisa Mitrofan for her thoughtful insights on the subject. All errors remain my own.

surrounding saviour siblings, commercial organ donation and commercial surrogacy, as well as considering the specific arguments against each. The article concludes that whilst the law strikes a right balance in the context of saviour siblings, it fails to do the same in the context of organ donation and surrogacy by failing to recognise the value of their commercial use. Crucially, I argue that the concerns regarding the commercialisation of organ donation and surrogacy are overstated and can be mitigated by establishing a regulatory framework similar to that which currently regulates the creation of saviour siblings.

Introduction

For some, advances in modern medicine have brought new hope to individuals and families struggling with life-threatening conditions and reproductive issues. For others, relatively new practices such as the creation of saviour siblings, surrogacy and organ transplants raise a set of serious ethical concerns regarding the welfare of those participating in these practices. Whether these concerns ought to outweigh the risks and disadvantages posed by the current restrictive regulation of these practices will be the subject matter of this article. The article's structure is as follows.

Section 1 first sets out the relevant law for the regulation of these practices. Section 2 will then address the preliminary question of why we seek to protect the body from commodification in the first place. In particular, it evaluates the arguments made by Margaret Radin on this subject and makes two counterarguments.¹ The first is that we must unpick our intuitions concerning what is morally impermissible; moral intuitions alone do not provide a stable basis for rejecting a practice. The second is that the claim that market discourse is antagonistic to the interests of personhood is overstated. Next, section 3 argues that the current law adequately deals with the more specific objections of commodification and harm in relation to the creation of a child whose tissue can be used to provide treatment for the condition of a seriously ill sibling. The emotive term 'saviour siblings' has come to be used in such cases and will be used throughout the article, though the practice is officially known as preimplantation tissue typing. Section 4 then engages with the three most salient arguments against legalising the practice of commercial organ donation and surrogacy. First, it

¹ Margaret Radin, 'Market-Inalienability' (1987) 100 HLR, 1849.

provides a brief evaluation of the concerns of commodification, specifically in the context of commercial organ donation and surrogacy. Second, it addresses fears of inequity and exploitation. Third, it considers whether commercialisation would detract from the practices' inherent value in fostering interpersonal relationships. Section 5 concludes that the current regulatory system for the creation of saviour siblings should be maintained and should form the template for regulating commercial organ donation and surrogacy.

1. The Current Law

A. Saviour Sibling Selection

The option for families to give birth to a saviour sibling was held to be legal by the seminal House of Lords case, *Quintavalle v Human Fertilisation and Embryology Authority (Quintavalle)*.² In general, assisted reproductive technologies (ARTs)³ are regulated by the Human Fertilisation and Embryology Act 1990 (HFE Act 1990), which created the Human Fertilisation and Embryology Authority (HFEA). The HFEA issues guidance on ARTs and licenses their use. Under section 3(1) of the HFE Act, it is a criminal offence to bring about the creation of an embryo except under licence from the Human Fertilisation and Embryology

² [2005] UKHL 28.

³ These include, *inter alia*, In Vitro Fertilisation, Intra-Uterine Insemination, In Vitro Maturation, Vitriification, and Intra-Cytoplasmic Sperm Injection.

Authority (HFEA).⁴ Although the medical authority in *Quintavalle* had been granted a licence by the HFEA, the legality of the licence was challenged by the group Comment on Reproductive Ethics (CORE). Its application succeeded before Maurice J, but failed before the Court of Appeal and the House of Lords, thereby clarifying that the HFEA can continue to issue licences for the creation of a saviour sibling.

In addition, the legality of the creation of saviour siblings was further clarified by the HFE Act 2008, which amended the 1990 Act. Pursuant to the 2008 Act, the creation of saviour siblings is permissible, subject to certain qualifications.⁵ For example, it can only be used in the case of siblings⁶ and paragraph 1ZA(1)(d) states that the creation of a saviour sibling is impermissible if it is planned that a whole organ is to be donated. Only bone marrow, umbilical cord blood stem cells, or other tissue may be used.⁷ Moreover, when a clinic is deciding the appropriateness of preimplantation tissue typing, it must consider

⁴ See also paragraphs [7]-[9] of Lord Hoffmann's judgment for full discussion of the history of the Act and details regarding its application.

⁵ Of particular importance is paragraph 1ZA(d) of Schedule 2 of the 1990 Act, which allows testing where a person ('the sibling') who is the child of the persons whose gametes are used to bring about the creation of the embryo (or of either of these persons) suffers from a serious medical condition which could be treated by umbilical cord blood stem cells, bone marrow or other tissue of any resulting child, establishing whether the tissue of any resulting child would be compatible with that of the sibling.

⁶ See paragraph 1ZA(1)(d) which is concerned with 'tissue typing' and sole reference to 'the sibling'.

⁷ For reasons of space, whether this distinction is justified is beyond the scope of this article.

the condition of the affected child,⁸ the possible consequences for any child who may be born as a result,⁹ and the family circumstances of the people seeking treatment.¹⁰

B. Commercial Organ Donation and Surrogacy

By contrast, while neither surrogacy nor organ donation is illegal, it is nevertheless an offence to arrange, negotiate or even offer to negotiate such exchanges on a commercial basis,¹¹ as is advertising such services.¹² However, the prohibitions on

⁸ Human Fertilisation and Embryology Authority Code of Practice 2018 (HFEA Code 2018) [10.25]: (a) the degree of suffering associated with their condition; (b) the speed of degeneration in progressive disorders; (c) the extent of any intellectual impairment; (d) their prognosis, considering all treatment options available; (e) the availability of alternative sources of tissue for treating them, now and in the future; and (f) the availability of effective therapy for them, now and in the future.

⁹ *ibid* [10.26]: (a) any possible risks associated with embryo biopsy; (b) the likely long-term emotional and psychological implications; (c) whether they are likely to require intrusive surgery as a result of the treatment of the affected child (and the likelihood that this will be repeated); and (d) any complications or predispositions associated with the tissue type to be selected.

¹⁰ *ibid* [10.9]: (a) the views of the people seeking treatment in relation to the condition to be voided, including their previous reproductive experience; (b) the likely degree of suffering associated with the condition; (c) the availability of effective therapy, now and in the future; (d) the speed of degeneration in progressive disorders; (e) the extent of any intellectual impairment; (f) the social support available; and (g) the family circumstances of the people seeking treatment.

¹¹ Surrogacy Arrangements Act 1985, s 2(1). Notably, neither the commissioning parents nor the surrogate can be guilty of this offence under s 2(2). The prohibition of commercial dealings in human material for transplantation is set out in the Human Tissue Act 2004, s 32(1).

¹² Surrogacy Arrangements Act 1985, s 3; Human Tissue Act, s 32(2).

payment for organs and surrogacy are not as strict as it might first appear. Crucially, in section 32(7) of the Human Tissue Act 2004, the prohibitions on reward do not apply to expenses incurred in removing, storing or transporting the donated organ. Nor do they include payments to cover the loss of earnings of a person donating an organ.¹³

Importantly, payments can be made to cover a surrogate's expenses. The courts have interpreted this provision widely, using it to give primacy to the welfare of the child whilst essentially overlooking dimensions of profit. The first reported case which illustrates this is *Re C (Application by Mr and Mrs X)*,¹⁴ which considered section 54(8) of the HFE Act 2008. In this domestic surrogacy case, the commissioning parents had been misled by the surrogate to believe that she would suffer loss of earnings although she had not previously been in employment. As such, parts of the payments made to the surrogate by the commissioning parents were not reimbursement of reasonably incurred expenses when caused by the pregnancy. Wall J determined that the court had the power to give retrospective authorisation to such payments.¹⁵ Citing Latey J's judgment in *Re Adoption Application (Payment for Adoption)*,¹⁶ His Lordship noted that it would be 'draconian' to prohibit adoption in all instances in which there is an element of profit.¹⁷ His Lordship accepted the test posited by Latey J, which balanced all the circumstances with the welfare of the child, as the 'correct one'.¹⁸ Thus, though Wall J gave weight to the reasonableness of the sum provided to the surrogate, and the commissioning parents' lack of knowledge

¹³ Human Tissue Act 2004, s 32(1).

¹⁴ [2002] EWHC 157 (Fam).

¹⁵ *ibid* [29]-[35].

¹⁶ [1987] Fam 81.

¹⁷ *Re C* (n 14) [29].

¹⁸ *ibid* [30].

regarding the surrogate mother's lack of employment (and therefore the lack of lost income), His Lordship stressed the genuine nature of the couple's intention and the love that the child would receive.¹⁹

Concern for the welfare of the child was elevated further by the importation of section 1 of the Adoption and Children Act 2002 into the court's determination of parental orders under s 54 of the HFE Act 2008. Notably, the courts have shown in *Re P-M*²⁰ that they are even more willing to authorise payments to organisations acting commercially, provided the child's welfare is ensured. *Re P-M* underlined that it would take the 'most exceptional case' to refuse a parental order.²¹ Here, Theis J clarified that the ambit of section 54(8) was not limited to payments made to the surrogate mother but extended to payments made to an organisation which is effectively acting on a commercial basis.²² Her Ladyship recognised that the sum paid went beyond reasonable expenses.²³ Yet, she also noted that what constitutes a reasonable sum for the facilitation of surrogacy arrangements is 'determined by what people are prepared to pay'.²⁴

In summary, the creation of saviour siblings is permissible, subject to certain qualifications, whereas payment beyond covering an organ donor's or surrogate's expenses is not. However, in the context of surrogacy, the courts have adopted a

¹⁹ *ibid* [34].

²⁰ [2013] EWHC 2328 (Fam).

²¹ *ibid* [19].

²² *ibid* [15].

²³ *ibid* [17].

²⁴ *ibid* [21]. This was confirmed in *Re C* (n 14) [34] in which Theis J held that notwithstanding an element of commercialisation, there was no moral taint because the child was living in a home environment where he was cherished and loved.

generous interpretation of what constitutes reasonable expenses and given priority to the welfare of the child in instances where the surrogate and surrogacy agency clearly profit. How widely the courts will interpret the prohibition on commercial exchanges in the context of organ selling is yet to be tested. The main arguments used to justify the law's restrictive approach to the sale of organs and surrogacy services are examined below.

2. Why do we Protect the Body from Commodification?

A. The Arguments

Margaret Radin's piece 'Market-Inalienability'²⁵ provides us with a set of arguments against the commodification of the body. She argues that market discourse²⁶ itself might be antagonistic to the interests of personhood, which she defines as a commitment to 'an ideal of individual uniqueness.'²⁷ This is because systematically conceiving of personal attributes as fungible objects is threatening to personhood, as it detaches that from a person which is integral to them. She maintains that bodily integrity is an essential attribute, not detachable from the body. Therefore, she reasons,

²⁵ Radin (n 1).

²⁶ By which I mean adopting the kind of rhetoric used when dealing with fungible goods (in a marketplace). The concern is that such rhetoric does harm to an ideal of uniqueness because of the equalising nature of ascribing monetary value to something. There is a concern that something becomes interchangeable, and is therefore no longer unique, once this kind of discourse is employed.

²⁷ Radin (n 1) 1885.

‘hypothetically valuing my bodily integrity in money is not far removed from valuing me in money.’²⁸ This, she concludes, is ‘intuitively wrong’ because it makes public what is otherwise personal and means we must accept an inferior conception of human flourishing by ‘reduc[ing] the conception of a person to an abstract, fungible unit with no individuating characteristics.’²⁹ This position, when considered in the specific context of justifying the permissibility of saviour siblings, organ selling and commercial surrogacy, can be challenged on two grounds, explored below.

B. Unpicking Our Intuitions

First, Radin’s reliance on what is ‘intuitively wrong’ needs to be addressed as she uses phrases such as ‘deep intuitive appeal’,³⁰ ‘intuitive grasp’,³¹ and appeals to our ‘aversion to commodification’.³² On the one hand, the British Medical Association (BMA) has suggested that intuition can be a useful component of ethical guidance, even if dangerous when used alone.³³ Indeed, it is unlikely that ethical guidance issued to medical professionals that goes against the dictates of their conscience would be respected or effective. On the other hand, sceptics, such as Ruth Macklin, claim that the perception of something as “yuck” is a conversation stopper, not an argument.³⁴ The way to reconcile these opposing views is to dig

²⁸ *ibid.*

²⁹ *ibid.*

³⁰ *ibid* 1926.

³¹ *ibid.*

³² *ibid* 1928.

³³ A Sommerville, ‘Juggling Law, Ethics, and Intuition: Practical Answers to Awkward Questions’ (2003) 29(5) JME 281.

³⁴ Ruth Macklin, ‘Can one do good medical ethics without principles?’ (2015) 41(1) JME 75.

deeper in order to see what underpins the intuitive aversion to these practices. This way, we can give weight to those feelings without them stifling the discussion.

Two arguments support Radin's claims of the existence of an intuitive aversion to commodification. The first is grounded in dignity, the significance of which principle is recognised by the Council of Europe's Convention on Human Rights and Biomedicine.³⁵ However, as Beyleveld and Brownsword have usefully noted, the principle can be used as both a means of empowerment and as a constraint.³⁶ Dignity as empowerment sees human dignity as a source of rights. It requires the respecting of individual choices such that they may live an autonomous life.³⁷ Dignity as a constraint refers to what a community regards as special about individuals. For example, Killmister highlights that whereas autonomy is concerned with self-government, dignity is concerned with self-worth.³⁸ Thus, even in the case of separated bodily material or deceased bodies, which can no longer be autonomous, the same principle of dignity should apply. For example, using a human head as a cereal bowl falls below the standard of how we would like to be treated and offends our

³⁵ Quoted in Jackson, *Medical Law and Ethics* (4th edn, OUP 2016) 25: 'to take such measures as are necessary to safeguard human dignity and the fundamental rights and freedoms of the individual with regard to the application of biology and medicine'.

³⁶ Beryck Beyleveld and Roger Brownsword, 'Human Dignity and the New Bioethics: Human Dignity as Constraint' in Beyleveld Deryck (ed), *Human Dignity in Bioethics and Biolaw* (OUP 2001).

³⁷ See Joseph Raz, 'The Rule of Law and its Virtue' in *The Authority of Law: Essays on Law and Morality* (Clarendon Press 1979) 221. He argues that 'respecting people's dignity includes respecting their autonomy'.

³⁸ Suzanne Killmister 'Dignity: not such a useless concept' (2010) 36 JME 160.

conception of self-worth, even when using it as such is not an infringement on autonomy per se.

Indeed, it is this latter conception of dignity that seems to capture Radin's views best. At various points in her article, she refers to 'an ideal of human flourishing'.³⁹ In fact, this notion of human flourishing is deemed by some, notably Foster, to be central to the principle of dignity.⁴⁰ However, Radin never defines this ideal. We can only find, in a footnote, an explanation that 'all we can do is reflect on what we know about human life and choose the best from among the conceptions available to us'.⁴¹ Importantly, she does recognise that 'we should not accept a conception of human flourishing that excludes our understanding of politics as (also) community self-determination, excludes our understanding of reproductive capacity as essentially human and personal'.⁴² This recognition of the importance of self-determination and reproductive capacity as 'essentially human and personal' is fundamentally at odds with her unqualified assertion that '[c]ommodification brings about an inferior form of human life'.⁴³ In fact, section 3 will argue that the permissibility of saviour siblings, commercial organ donation and commercial surrogacy *can* respect the self-determination and reproductive autonomy of its community. Ultimately, as Cochrane has argued, dignity as constraint is used by those who want to stop people who they believe are behaving immorally.⁴⁴

³⁹ Radin (n 1) 1849, 1903, 1923.

⁴⁰ Charles Foster, *Dignity in Bioethics and Law* (Hart Publishing 2011) 6.

⁴¹ Radin (n 1) 1884.

⁴² *ibid.*

⁴³ *ibid.* See also pages 1872, 1886, 1908, 1912 and 1927.

⁴⁴ Alasdair Cochrane, 'Undignified Bioethics' (2010) *Bioethics* 24(5) 234.

The second argument is rooted in the generally accepted principle of the sacredness of life. From a religious perspective, there are members of religious groups that think that humankind was created in God's image and, therefore, that the body is inherently sacred.⁴⁵ Although full discussion of the religious dimension of this principle is beyond the scope of this article,⁴⁶ it is important to acknowledge its existence in light of how widely held the view is.

There is also a secular basis for this argument.⁴⁷ In chapter three of Ronald Dworkin's *Life's Dominion*, he observes

⁴⁵ In the Hebrew Bible, at Genesis 1:26-28: 'And God created man in His image'. Genesis 9:6: 'One who spills the blood of man, through/by man, his blood will be spilled, for in God's image he made man.' In the New Testament, see 1 Corinthians 11:7: 'For a man ought not to have his head covered, since he is in God's image and glory'. Although the Quran stresses that nothing is nothing is akin to God (Surah Ash Shura, verse 11), it echoes the other two Abrahamic religions in noting that God created Adam in his own image (Sahih Bukhari Hadith n 6227). Outside of the three Abrahamic religions, in the Buddhist tradition, the body and mind are believed to be mutually dependent. The body or physical form is considered one of the five skandhas, the five interdependent components that constitute an individual—Liz Wilson, 'Perspectives on the Body', in Robert Buswell Jr (ed), *Macmillan Encyclopaedia of Buddhism* (New York: Macmillan Reference 2004) 63.

⁴⁶ For reasons of space and due to the spectrum of views within different religious groups that I cannot possibly hope to accurately represent. Further, there is a sound argument against basing the law on religious views in light of the multicultural nature of society in the UK. Giving too much weight to the religious views of one group may result in majoritarian, or even minoritarian, rule over those who do not share these views.

⁴⁷ See also Margaret Somerville, *Death Talk* (McGill-Queen's University Press 2001) who also emphasises the 'secular sacred'. In particular, she stresses the 'deeply intuitive sense of relatedness or connectedness to other people and to the world and universe in which we live'. Again,

that there is a general consensus of regret when human life, once begun, ends prematurely.⁴⁸ In addition to noting that the idea of *intrinsic* value is familiar in our shared scheme of values,⁴⁹ he attempts to justify this claim by reference to his conception of the sacredness of life. Dworkin shows that our varying senses of the sacred or inviolate lies either in association with some other value, like a flag with its country, or because of the human, natural or divine process that brought the sacred object or entity into being. The essence ‘of the sacred lies in the value that we attach to a process or enterprise or project rather than to its results considered independently from how they were produced’.⁵⁰ Our reverence towards our bodies, then, extends from the value we ascribe to life. This is because our bodies constitute the boundary and the expressive link between our subjective and objective world, as well as the site where all the processes of life happen.

However, we can accept the notion that the body is a vessel for personhood and still also recognise that the body’s capacity to be such a vessel is not threatened when we use or modify specific parts of the body. What makes us valuable is not affected by our choice to use our body instrumentally or to modify it. This is made clear by the fact that we do not only consider whole, unmodified bodies to be valuable or the only suitable vessels of personhood. The aversion towards commercialising organ selling, such as kidneys or parts of the liver (where their removal does not pose an imminent threat to one’s

see also: Somerville, ‘Deathbed disputation’ (2002) 167 CMAJ 651, 654.

⁴⁸ Ronald Dworkin, *Life’s Dominion* (Vintage 1994) 86.

⁴⁹ *ibid* 90-94. Dworkin points to a collective sense that much of what we think about knowledge, experience, art and nature presupposes that in various ways each is valuable in themselves and not just for their utility or for the pleasure or satisfaction that they bring us.

⁵⁰ *ibid* 97.

health) and surrogacy are a result of society's taboos as opposed to an objective difference between such practices and other, accepted practices. For example, undergoing some types of cosmetic surgery poses similar health risks to those posed by organ transplants.⁵¹ In fact, in 52 studies comprising 118,426 living kidney donors and 117,656 nondonors, there was no evidence suggesting higher risk for all-cause mortality, cardiovascular disease, hypertension, type 2 diabetes, or adverse psychosocial health outcomes in living kidney donors than in nondonor populations.⁵² Such evidence accounts for long-term risks too, as the average follow-up in these studies was 1 to 24 years.⁵³

Moreover, the law does not only permit minor cosmetic alterations, such as laser hair removal or lip fillers. The law permits *radical changes* to the body, such as breast implants, pulling out one's teeth and even piercings in parts of the body that are full of nerve endings that can be damaged as a consequence.⁵⁴

⁵¹ Giuliano Testa, Erica Carlisle, Mary Simmerling, and Peter Angelos, 'Living Donation and Cosmetic Surgery: A Double Standard in Medical Ethics?' (2012) 23(2) JCE 110 <<https://pubmed.ncbi.nlm.nih.gov/22822698/>> accessed 20 November 2020.

⁵² Linda O'Keeffe, Anna Ramond, Clare Oliver-Williams, et al., 'Mid- and Long-Term Health Risks in Living Kidney Donors: A Systematic Review and Meta-analysis' (2018) 168(4) *Annals of Internal Medicine* 276 <<https://pubmed.ncbi.nlm.nih.gov/29379948/>> accessed 19 March 2021.

⁵³ *ibid.*

⁵⁴ For example, tongue piercings are frequently accompanied by a small amount of nerve damage, even if the procedure has been properly carried out. Notwithstanding this risk, the patient will be informed of these risks and is free to go ahead with the procedure. Dirk Ziebolz, Else Hornecker, Constatin Von See and Rainer Mausberg, 'Complications of Tongue Piercing: A Review of the Literature and Three Case Reports' (2009) 10(6) JCDP.

Similarly, the procedure of Botox is carried out by injecting an individual with the bacterium *Clostridium botulinum*, which is the same toxin that causes life-threatening food poisoning *botulism*.⁵⁵ In small doses, however, its use is considered safe.⁵⁶ Thus, we can see that the law also permits individuals to pay for procedures which carry risks that are separate and additional to those relating to a faulty operation. Most importantly, the law permits such practices because the risks are sufficiently low as to be outweighed by other salient considerations.

In all the aforementioned examples, the intactness of the body is compromised and yet, in the case of plastic surgery, the law does not allow an intuitive aversion to dictate what is or is not permissible. Indeed, the perception that an organ such as a liver (which is regenerative) is substantially different than skin is misconstrued. After all, the skin is the body's largest organ.⁵⁷ The additional argument that putting a monetary value on an organ donation is dehumanising is considered and rejected below.

Finally, in the case of surrogacy, we must remember that we permit individuals to undergo dangerous construction work for the sake of payment. Such jobs often pay a premium because the integrity of the body is at risk. This risk arises from the possibility of negligence, as a momentary lapse of concentration can result in disastrous consequences. However, such work is also

⁵⁵ 'Botulism' (NHS UK, 2018)
<<https://www.nhs.uk/conditions/botulism/>> accessed 12 November 2020.

⁵⁶ Timothy Coté, Aparna Mohan, Jacquelyn Polder, et al., 'Botulinum Toxin Type A Injections: Adverse Events Reported to the US Food and Drug Administration in Therapeutic and Cosmetic Cases' (2005) 53(3) JAAD 407.

⁵⁷ 'Skin And How It Functions' (*Science*, 2017)
<<https://www.nationalgeographic.com/science/health-and-human-body/human-body/skin/>> accessed 5 February 2021.

inherently harmful, notwithstanding the absence of negligence, because it invariably takes place in very dirty, dusty and polluted environments. For example, in *Wilsher v Essex Area Health Authority*,⁵⁸ Lord Bridge distinguished between the ‘innocent’ and ‘guilty’ dust particles that the claimants in *Bonnington Castings Ltd v Wardlaw*⁵⁹ and *McGhee v National Coal Board*⁶⁰ were exposed to. Although such a distinction can help the court in deciding whether or not the element of causation is present in a negligence claim, the distinction rests on a fiction. Both dust particles are guilty of adversely affecting a worker. Thus, in both the case of construction work and commercial surrogacy, the sanctity of the body has been compromised and there is a commercial dimension. Yet, only the latter is, according to statute, legally impermissible.⁶¹

In fact, the Health and Safety Executive estimate that around four percent of construction workers suffer from a work-related illness every year, and 3% sustain a work-related injury.⁶² The total figure of 7% risk of injury is analogous to the 8% of pregnancies that involve complications that, if left untreated, may harm the mother or the baby.⁶³ In both instances, decisions over our body are an expression of our personhood which, as Radin notes, concerns ‘individual uniqueness’. Therefore, to

⁵⁸ [1988] 1 All ER 871 at 1087H and 1088A-B.

⁵⁹ [1956] AC 613.

⁶⁰ [1973] 1 WLR 1.

⁶¹ Surrogacy Arrangements Act 1985, s 2(1).

⁶² ‘Injuries on Construction Sites’ (*Designing Buildings Wiki*, 10 May 2021)

<https://www.designingbuildings.co.uk/wiki/Injuries_on_construction_sites> accessed 14 March 2021.

⁶³ ‘Four Common Pregnancy Complications’ (*John Hopkins Medicine*, 2001) <<https://www.hopkinsmedicine.org/health/conditions-and-diseases/staying-healthy-during-pregnancy/4-common-pregnancy-complications>> accessed 14 March 2021.

paternalistically disallow the commercialisation of surrogacy and organ donation due to an intuitive aversion would render the law arbitrary. Given the similarities between the practices, such interventions should be guided by the same ethical concern for one's autonomy. In summary, then, the 'intuition' Radin discusses is an inadequate basis upon which to reject the partial commodification of the body.

C. Market Discourse as Antagonistic to Personhood

Second, and relatedly, the concern that market discourse is in itself antagonistic to the interests of personhood, which Radin equates with human flourishing, is overstated. In order to substantiate this, we can take a closer look at what she understands personhood to mean. She unpacks the notion of 'individual uniqueness' with examples of one's 'politics, work, religion, family, love, sexuality, friendship, altruism, experiences, wisdom, moral commitments, character, and personal attributes'⁶⁴ as integral to oneself. Monetising or completely detaching any of these from the person, she argues, is to do 'violence to our deepest understanding of what it is to be human'.⁶⁵

There are two counterarguments that could be made in response to this. We can contest the claim that the monetisation of the aforementioned examples would necessarily cause such 'violence' to one's personhood. For example, her reference to 'work' stands in sharp contrast to the fact that the vast majority of individuals undertake work for money and are able to separate

⁶⁴ Radin (n 1) 1906.

⁶⁵ *ibid.*

their work from the rest of their lives.⁶⁶ Similarly, many life experiences, such as a trip to another country, unavoidably can be monetised. This does not mean that the work or experience should be reduced to that sum. Rather, it illustrates that we can monetise something without it being overly disruptive to its inherent value and, by extension, our personhood.

Admittedly, a distinction could be drawn between extrinsic value such as love, friendship and wisdom and the physical body itself, without which we cannot live. Yet, such a distinction would be uninformative unless you value all aspects of the body as essential to our existence, such that it would be wrong to cut our hair or nails, or undergo any cosmetic procedure. Ultimately, both are a matter of degree. Absolute monetisation of the body or extrinsic values such as love would harm one's personhood, as the person would be dehumanised, rendered non-existent except by reference to their monetary value. It is for this reason that this article does not propose such an argument. Rather, it seeks to find the precise point at which harm can arise in the specific context of saviour siblings, commercial organ donation and commercial surrogacy, as outlined in section 3.

Moreover, a stronger basis for rejecting Radin's argument is to prove that the specific issues of saviour siblings, commercial organ donation and surrogacy may be facilitative of the particulars that constitute personhood. For example, as Spriggs has noted, the ability to bring a saviour sibling into existence, to save a sibling, will benefit both Max (the saviour sibling) and Max's family by enabling Max to be born into a family

⁶⁶ Indeed, Radin (n 1) 1914 recognises as much when she notes that 'many people value their homes or their work in a nonmonetary way, even though things also have market value'.

which is not wrecked by bereavement.⁶⁷ Thus, the family unit is protected, enabling the individuals within that unit and the collective family to flourish. The counterarguments to this position, such as the potential harms to the child, are considered and rejected below.

Similarly, commercial surrogacy may be facilitative for greater human flourishing, as friendships, wonderful life experiences and a family full of love may be created as a consequence. In *Re P-M*, Theis J, observed that notwithstanding a clear element of profit, the prospective parents and the surrogate had maintained a positive relationship.⁶⁸ In fact, the prospective parents were willing to cover the surrogate's travel costs such that the children could meet their surrogate mother.⁶⁹ A similar sentiment is expressed in *Re C*.⁷⁰ Furthermore, in both cases, there can be no doubt that the children and the surrogate were treated with 'tenderness',⁷¹ or were 'loved and cherished'.⁷²

Moreover, the National Institute for Health and Care Excellence (NICE) estimate that infertility affects 1 in 7 heterosexual couples.⁷³ Such a figure may only increase in the next decade as couples delay having children until much later in their lives. Furthermore, the ability to conceive a child is even less straightforward for same-sex couples. The inability to have

⁶⁷ Merle Spriggs, 'Is conceiving a child to benefit another against the interests of the new child?' (2005) 31(6) JME 341.

⁶⁸ *Re P-M* (n 20) [22].

⁶⁹ *ibid* [34].

⁷⁰ *Re C* (n 14) [18].

⁷¹ *Re P-M* (n 20) [25].

⁷² *Re C* (n 14) [34].

⁷³ National Institute for Health and Care Excellence, 'Fertility Problems Briefing Paper' (2014) [2.3] <<https://www.nice.org.uk/guidance/qs73/documents/fertility-problems-briefing-paper2>> accessed 11 December 2021.

control over one's reproductive capacity is linked to significant anxiety and emotional distress.⁷⁴ Yet, commercial surrogacy provides a viable solution for these couples.

Admittedly, the underlying cause of infertility has not been remedied. Further, why should the suffering of one couple be used to justify permitting the commercialisation of someone else's reproductive capacity? Due to the shortage of surrogates in the UK, almost two-thirds of all UK parental orders⁷⁵ are for a baby born overseas.⁷⁶ Introducing the commercial dimension would incentivise more women to become surrogates and to be properly remunerated as a consequence.⁷⁷ Although covering a surrogate's reasonable expenses is permitted,⁷⁸ the option of becoming a surrogate remains prohibitively expensive. This is because many young women understandably wish to *increase* their aggregate wealth. Of course, pregnant women can continue to work. Yet, the reality is that pregnancy does impact whether they are likely to be hired for many jobs and how long they will be part of the workforce before taking leave.⁷⁹ Ultimately, permitting a

⁷⁴ 'Infertility and Mental Health' (*Cedars-Sinai Blog*, 8 September 2020) <<https://www.cedars-sinai.org/blog/infertility-mental-health.html>> accessed 15 March 2021.

⁷⁵ Which means a legal right conferred on parents who have commissioned a child from a surrogate.

⁷⁶ Jamie Doward, 'Childless UK Couples Forced Abroad to Find Surrogates' *The Guardian* (London, 20 February 2016) <<https://www.theguardian.com/lifeandstyle/2016/feb/20/childless-uk-couples-forced-abroad-surrogates>> accessed August 2020.

⁷⁷ The argument that this will further reduce the risks of exploitation abroad is considered in more detail below.

⁷⁸ See section 1 of this article, 224-228.

⁷⁹ For example, although the Equality Act 2010 prohibits the discrimination on grounds of protected characteristics, such as pregnancy and maternity, only *reasonable adjustments* need to be made. It would therefore not be unlawful to not hire an individual where there

commercial dimension would enable commissioning parents, as well as surrogates, to have self-determination by being free to decide an essentially personal matter concerning their reproductive capacity.⁸⁰

Finally, it is plausible to permit the selling of non-vital organs, such as kidneys or parts of the liver (where their removal does not pose an imminent threat to one's health) without harming our notion that the body is a space of personhood that should be protected. Whilst in many instances such an act would not threaten either our survival or holistic value as a person, it could be necessary for the survival of another. In fact, there are around 6,000 people on the UK Transplant Waiting List, and, last year, over 350 people died while waiting for a transplant.⁸¹ Moreover, Radin's assertion that it is damaging to detach from the person that which is integral to them presupposes that such body parts are in fact integral to them.⁸² This is inconsistent with the biological reality of the importance of certain organs to the

are concerns that they would not be able to cope with the physical demands of the position where there are equally well-qualified candidates who are likely to be more physically capable of performing the role. It would also not be reasonable for a two-partner firm to take any measure to accommodate a pregnancy where there are other equally-qualified candidates.

⁸⁰ The possible counterargument regarding the disparity in bargaining power between the surrogate and commissioning parents is discussed below.

⁸¹ 'Organ Donation and Transplantation' (*NHS Blood and Transplant*, 2021) <<https://www.nhsbt.nhs.uk/what-we-do/transplantation-services/organ-donation-and-transplantation/>> 15 January 2021.

⁸² Radin (n 1) 1885. Admittedly, Radin makes this point in the context of surrogacy and does not address organ donation. Nonetheless, the argument made against surrogacy (notably against even non-commercial surrogacy) can be applied to the context of commercial organ donation.

wellbeing of the donor. For example, it is well documented that a kidney donor can lead a normal life with only one functioning kidney.⁸³ More importantly, it illiberally superimposes the view that such body parts are integral to personhood,⁸⁴ thereby paternalistically undermining the autonomy of the donor. Finally, as noted above, it is incongruent with Radin's own conception of human flourishing, which, as she observes, must give weight to self-determination.

On the other hand, Kimel, adopting a similar perspective to that espoused by Green over a hundred years earlier,⁸⁵ argues that concern for personal autonomy does not entail the extreme conception often associated with liberalism, which entails unqualified freedom.⁸⁶ Central to his argument is the view that state intervention is positively required to ensure that, 'by and large, this freedom would enhance the well-being of those who enjoy it, and make a positive contribution to their chances of

⁸³ 'Become A Living Donor' (*NHS Organ Donation*, 2021) <<https://www.organdonation.nhs.uk/become-a-living-donor/>> accessed 15 January 2021.

⁸⁴ For the purposes of this point, a conception of 'liberalism' is adopted which accords generally with the following definition of 'deontological liberalism' by Michael Sandel, *Liberalism and the Limits of Justice* (Cambridge University Press 1998) 1: '[L]iberalism is above all a theory about justice, and in particular about the primacy of justice among moral and political ideals. Its core thesis can be stated as follows: society, being comprised of a plurality of persons, each with his own aims, interests, and conceptions of the good, is best arranged when it is governed by principles that do not themselves presuppose any particular conceptions of the good.'

⁸⁵ Thomas Green, 'Liberal legislation and Freedom of Contract' in *Lectures on the Principles of Political Obligation and Other Writing* (CUP 1986) 194.

⁸⁶ Dori Kimel, 'Neutrality, Autonomy, and Freedom of Contract', 21(3) *Oxford Journal of Legal Studies* 473.

leading valuable, successful lives.⁸⁷ The crux of his concern lies with the imbalance of power between contracting parties. This, he argues, undermines the claim that there is in fact freedom within a society which adopts a rigid conception of ‘freedom of contract’.

Yet, it is completely possible to propose a regulatory framework with safeguards against wrongful exploitation, which will be outlined further in section 4. Indeed, this article accepts Kimel’s criticism of the atomistic version of liberalism that has emerged as the prevailing attitude of our society. Such a concept is even more out of place in the context of medical law, whereby rights, welfare and social considerations are necessarily in balance with one another. However, prohibiting these practices in their entirety is the equivalent of using a sledgehammer when a scalpel is needed. As Radin herself conceded, ‘in attempting to make the hard choices in which both commodification and decommodification seem harmful’, what she terms the ‘double bind’, ‘we must evaluate each contested commodification in its temporal and social context’.⁸⁸

In fact, the courts clearly had this idea in mind in *Re P-M*.⁸⁹ Despite section 54(8) of the HFE Act 2008 explicitly proscribing payments beyond covering mere expenses, judges have found it necessary to interpret ‘expenses’ widely in order to protect the best interest of the parties involved. This, it is submitted, has been the right approach for the courts to take. Rather than adhering to the argument that commercialising surrogacy leads to a devaluation of the parties that the court cannot condone, it has taken an approach akin to that which

⁸⁷ *ibid* 494.

⁸⁸ Radin (n 1) 1915-1917, 1937.

⁸⁹ *Re P-M* (n 20).

Radin proposed.⁹⁰ However, this has led to a hollowing of the statute, as the courts have explicitly tolerated payment which exceeds the amount needed to cover expenses to not only the surrogate,⁹¹ but to international agencies too.⁹² Further, there is much uncertainty regarding how ‘expenses’ will be interpreted in the context of organ donation. The proper course to take at this juncture is establish a regulatory framework which is able to evaluate each contested commodification in its temporal and societal context.

Ultimately, willingly using one’s body instrumentally can be both conducive to individual and collective human flourishing. As Green points out, being needed instrumentally can be an important part of human fulfilment; we desire to feel useful as means as well as ends in ourselves.⁹³ It only becomes morally impermissible where we reduce a person or their body to their instrumental value to the extent to which we ‘drown out the[ir] subjective worth’—in other words, when we see bodies as *merely* of instrumental worth.⁹⁴ The difficulty for the law is in finding that balance.

⁹⁰ That ‘we must evaluate each contested commodification in its temporal and social context’, Radin (n 1) 1937.

⁹¹ *Re C* (n 14).

⁹² *Re P-M* (n 20).

⁹³ Leslie Green, ‘Pornographies’ (2000) 8(1) JPP 27.

⁹⁴ Kate Greasley ‘Property Rights in the Human Body: Commodification and Objectification’ in Imogen Goold, Jonathan Herring, Loane Skene, and Kate Greasley (eds), *Persons, Parts and Property: How Should We Regulate Human Tissue in the 21st Century?* (Hart Publishing 2014), 70.

4. Finding the Right Balance in the Context of Saviour Siblings, Commercial Organ Donation and Commercial Surrogacy

A. Saviour Siblings

One example of an area where the law strikes the right balance is in the context of saviour siblings. As noted in section 1, following the House of Lords judgment in *Quintavalle*⁹⁵ and the HFE Act 2008, embryo testing to ensure that an implanted embryo may be a saviour sibling is permitted.

I. Commodification

A major concern associated with the creation of saviour siblings is that it commodifies the child to be born⁹⁶ Underpinning this

⁹⁵ *Quintavalle* (n 2).

⁹⁶ Commodification is a multifaceted concept. I define this term broadly in my article to mean the action or process of treating something as an abstract, fungible unit with no individuating characteristics. The distinctive concern regarding the commodification of saviour siblings lies in treating something non-fungible and intrinsically valuable as though it were fungible and instrumentally valuable. In the case of saviour siblings, part of the child's body, often the bone marrow, is being treated as a fungible good to be used for the benefit of their sibling. A narrower definition might focus exclusively on whether the bodily material can be bought and sold. However, I would define this as 'commercialisation'. It is therefore possible to commodify the body without commercialising it, though the two rarely diverge in practice. I adopt this broad definition because there is ample scholarship which raises the issue of commodification of saviour siblings, notwithstanding the lack a commercial element to the exchange. For example, see Susan Wolf, Jeffrey Kahn, and John

objection is the deontological protestation against commodification that stems from Immanuel Kant's second formulation of the categorical imperative to treat people 'never simply as a means, but always at the same time as an end'.⁹⁷ The Kantian injunction against treating people simply as a means perceives the saviour child as having been wronged if the principle is offended, even if no harm flows to the child as a result.⁹⁸

However, policy makers in the UK have already addressed this concern. In 2001, the UK HFEA Ethics Committee stated that 'positive considerations of the welfare of the child requires respect for beings as ends and that the child in question be treated not simply as a means to further an end but also as an "end in itself"'.⁹⁹ Arguably, the UK could render this enquiry more explicit, and follow the approach taken by the National Health and Medical Research Council (NHMRC) in Australia. According to the NHMRC guidelines, ethics committees advising clinics on saviour sibling selection must ascertain whether the parents desire another child in his/her own right and not merely as a tissue source.¹⁰⁰ Nonetheless, we can find similar instructions in the UK with the HFEA Ethics

Wager, 'Using Preimplantation Genetic Diagnosis to Create a Stem Cell Donor: Issues, Guidelines and Limits' (2003) 31(3) JLME 327.

⁹⁷ Immanuel Kant, *Groundwork of the Metaphysics of Morals* (Revised Edition, Cambridge University Press 2005) [4.429].

⁹⁸ Colin Gavanagh, *Defending the Genetic Supermarket: Law and Ethics of Selecting the Next Generation* (Routledge-Cavendish 2007) 157: 'if all parties involved are net beneficiaries in terms of harms and benefits, we may have done something ethically wrong if, in the process, we treated some of them as mere instruments.'

⁹⁹ HFEA Ethics & Law Committee, *Ethical Issues in the Creation and Selection of Preimplantation Embryos to Produce Tissue Donors* (22 November 2001) [2.9].

¹⁰⁰ NHMRC, *Ethical Guidance on the Use of Assisted Reproductive Technology in Clinical Practice and Research* (June 2007) [12.3].

Committee's guidelines. It is not clear that it is necessary to formalise this instruction further.

Moreover, the objection is based on an incorrect interpretation of Kant's categorical imperative. As highlighted by Beauchamp and Childress, Kant's dictum prohibits treating another person *exclusively* as a means to an end.¹⁰¹ It is therefore acceptable to use someone as a means, provided you do not lose sight of the fact that they are also an end in themselves. After all, everyday transactions between employer and employee, client and service provider, and customer and vendor all involve treating someone as a means to an end and are not considered morally objectionable. Therefore, provided that the parents desire a child in their own right, even though they may also desire a child for a particular purpose, such as saving the life of an existing child, this objection can be easily met, and one can find that the law already strikes the right balance.

II. Harm

In contrast to the deontological argument regarding commodification, the harm objection focuses on the potential consequences for the saviour sibling. Any potential harm done to the child would contravene the keystone medico-legal principle of non-maleficence.¹⁰² This principle asserts an obligation not to inflict harm intentionally.

Yet, the UK's HFEA policy reflects a 'harm-based' approach to the welfare of the child. Notably, there is already a

¹⁰¹ Tom Beauchamp and James Childress, *Principles of Biomedical Ethics* (5th edn, OUP 2001) 351.

¹⁰² Notably, it is one of Beauchamp and Childress' four core ethical principles in medical ethics: *ibid.*

threshold level of protection for the child.¹⁰³ In addition, following its review of the welfare of the child principle in 2005, the HFEA established a ‘presumption to provide treatment unless there is evidence that any child born to an individual or couple, or any existing child of their family, would face a risk of serious harm’.¹⁰⁴ This is reinforced by the HFEA’s Code of Practice, which requires assessments about the welfare of the child to be born to focus on factors that are likely to cause a risk of significant harm or neglect to the child.¹⁰⁵ As a consequence, clinics are required to refuse treatment where a risk of significant harm of neglect exists.¹⁰⁶ As such, this objection can also be easily met, and one can find that the law already strikes the right balance.

B. Commercial Organ Donation and Surrogacy

Given that the law accepts that it is morally permissible to create and use a human being instrumentally for the purposes of saving the life of another, why should that not extend to individuals choosing to use their *own* bodies instrumentally within the context

¹⁰³ In accordance with the ‘minimum threshold principle’, which is consistent with the House of Commons Science and Technology Committee (STC)’s recommendation: HFEA, *Tomorrow’s Children: A consultation on Guidance to Licensed Fertility Clinics on Taking in Account the Welfare of Children to be Born of Assisted Conception Treatment* (January 2005), [2.4]; STC, Fifth Report of Session 2004-5, *Human Reproductive Technologies and the Law*, HC7-1 (24 March 2005) (STC Report) [107].

¹⁰⁴ HFEA, *Tomorrow’s Children: Report of the Policy Review of Welfare of the Child Assessments in Licensed Assisted Conception Clinics* (January 2005), 1.

¹⁰⁵ HFEA, *Code of Practice* (8th edn, October 2011) (HFEA, Code of Practice) [8.15] <<https://www.hfea.gov.uk/media/2062/2017-10-02-code-of-practice-8th-edition-full-version-11th-revision-final-clean.pdf>> accessed 15 January 2021

¹⁰⁶ *ibid* [8.15].

of commercial surrogacy and organ selling? Three arguments are considered and rejected.

I. Commodification

In addition to reiterating the point that partial commodification does not harm an individual's personhood, it is possible to make a bolder claim. The concern with protecting personhood, while people's lives continue to be put at risk by the dearth of organs available for transplantation, is closing the viable option of establishing a market in organs from living donors. The reality is that cadaveric organs will never satisfy the growing demand for organs. Worldwide, hundreds of thousands, if not millions, die whilst waiting for a transplant.¹⁰⁷ As such, at the very least, we are facing a global crisis; at worse, we are facing a 'scandal',¹⁰⁸ due in large measure to the shortage of donor organs. Therefore, the debate ought to be framed in such a way as to establish a *prima facie* case for allowing organ sales. After all, even if we were to concede that some aspect of personhood is compromised by commoditising those organs that are sold, this must surely be outweighed by the possibility of saving lives and permitting vendors to choose for themselves what constitutes their best interests.

We must also recognise that sentiments, if we are to give social taboos any weight at all, can change. In the case of organ

¹⁰⁷ Charles Erin and John Harris, 'An Ethical Market in Human Organs' (2003) 29(3) *JME* 137. They estimate that there are around 700,000 patients on dialysis. In India alone, they estimate that 100,000 new patients present with kidney failure, and in Western Europe, 40,000 patients await a kidney but only 10,000 kidneys become available. Importantly, they emphasise that the figures are likely to be much higher as we do not know how many people fail to make it onto the waiting lists and so disappear from the statistics.

¹⁰⁸ *ibid.*

donation of the deceased, the change to an opt-out system of organ donation¹⁰⁹ captures a shift in our culture towards recognising that organ donation is, as Herring and Chau argue, not an ‘unnatural activity requiring an unusual degree of consent’, but rather ‘a natural part of the interaction between bodies.’¹¹⁰ Admittedly, it is possible to challenge Herring and Chau’s view on the basis that substituting a failing organ with foreign tissue is, by its very nature, unnatural. Indeed, the very practice of medical science, in prolonging life, is in this respect unnatural.

The issue becomes more complicated in the context of surrogacy due to the lack of separation between what has been commoditised (the surrogate’s uterus) and the individual (the surrogate). However, the nebulous harm done to human flourishing by this practice, as suggested by Radin, is an assertion without empirical grounding. Rather, the claim that it might be ‘degrading for the surrogate to commodify her gestational services’¹¹¹ does little more than maintain the very stigma that ought to be overturned. After all, in the instances in which this article would legalise commercial surrogacy, the individual has utilised their body in the way they wished to the degree that they consented. Moreover, it is not clear how most jobs that ‘poor’,¹¹² ‘ignorant’¹¹³ women would have to do, such as being a factory worker for the creation of useless materialistic objects, are more

¹⁰⁹ ‘Organ Donation Law in England’ (*NHS UK*, 2020)

<<https://www.organdonation.nhs.uk/uk-laws/organ-donation-law-in-england/>> accessed 24 March 2020. This took effect on 20 May 2020 and presumes the consent of the deceased individual in absence of express intention.

¹¹⁰ Jonathan Herring and Phong Chau, ‘My Body, Your Body and Our Bodies’ (2007) 15 *MLR* 34.

¹¹¹ Radin (n 1) 1930.

¹¹² *ibid* 1936.

¹¹³ *ibid* 1930.

conducive to human flourishing. Such a view illiberally presupposes one objective notion of the good and denies the individual from choosing it for themselves, a question which they are ultimately best situated to decide. Most importantly, it also fails to recognise that a surrogate may have more time to pursue her own interests—be it reading, painting, learning an instrument or a language—than if she were to work for an employer who would set the agenda and limit her free time. Ultimately, such practices may be conducive to a life the relevant person deems valuable because it gives them options to take on other hobbies, spend time with loved ones, save another's life, or give life to a family.

II. Inequity and Fear of Exploitation

A second argument put forward is that body parts should not be traded in markets because those selling their body parts will be subject to inequities and abuses. The concern is, as Murray argues, that these would have 'special significance in body markets, because it is the morally significant body that is being traded off.'¹¹⁴

The first example of this inequity and abuse is the risk that the rich will exploit the poor whose apparent choice to sell is not genuine because they are coerced by their economic circumstances. However, as Radcliffe-Richards has argued, the fact that poor people are denied the opportunity to sell their kidneys weakens rather than strengthens their position.¹¹⁵ A similar argument is made by Goold in the context of commercial

¹¹⁴ Thomas Murray, 'On the Human Body as Property: The Meaning of Embodiment, Markets, and the Meaning of Strangers' (1987) 20 *UMJLR* 1055, 1088.

¹¹⁵ Janet Radcliffe-Richards, Abdallah Daar, Raymond Hoffenberg, et al., 'The Case for Allowing Kidney Sales' (1998) 351 *The Lancet* 1950.

surrogacy.¹¹⁶ Crucially, both Goold and Radcliffe-Richards posit that realising our body's instrumental potential may be a powerful means of improving our circumstances.

Furthermore, addressing the underlying inequities that might lead someone to choose to commercialise their reproductive and non-reproductive organs as well as legalising these practices are not mutually exclusive. It would, as Radin herself concedes, be 'hypocritical'¹¹⁷ to deny individuals these options available to them until a large-scale redistribution of wealth and power occurs. Notably, we do not limit which jobs people can do due the constraints placed on them by their circumstances. Most importantly, this consideration must be weighed against other factors, such as the shortage of organs leading to death and the centrality of respecting a person's reproductive autonomy to their wellbeing.

A further concern, outlined but ultimately not supported by Arneson, is that surrogacy reinforces the ideological view that a woman's instrumental value lies in her role as a child bearer and domestic labourer.¹¹⁸ On this point, Raymond argues that surrogacy 'reinforces the perception and use of women as a breeder class'.¹¹⁹ Commercialising it would further reinforce these perceptions. Similarly, Radin takes this point even further, arguing that 'an oppressive understanding' of the interaction of surrogacy

¹¹⁶ Imogen Goold, 'Surrogacy: Is There a Case for Legal Prohibition?' (2004) 12(2) JLM 205, 211.

¹¹⁷ Radin (n 1) 1911.

¹¹⁸ Richard Arneson, 'Commodification and commercial surrogacy' (1992) 21(2) PPA 132, 162-63.

¹¹⁹ Janice Raymond, 'Reproductive Gifts and Gift Giving: The Altruistic Woman' (1990) 20(6) HCR 7, 11.

is ‘the most plausible one’, whereby ‘women are fungible baby-makers for men whose seed must be carried on’.¹²⁰

The answer to this concern, however, is education. It would be absurd to limit the choices that a person has, for example which career they can choose, on the basis that certain racial stereotypes might be reinforced. Similarly, it would be disproportionate to prohibit *all* women from choosing to be a surrogate, even where there is a financial gain, because certain sexist stereotypes might be reinforced. Counter-intuitively, this gives more power to the stereotypes and affects those who ought to be protected from those stereotypes, rather than those who hold these discriminatory views.

Indeed, Radin consistently highlights the power of rhetoric in shaping our perceptions. Yet, there is no attempt to use language which might lessen the stigma for those women who might be proud to help the creation of a family, nor in the case where one might choose to be a surrogate for financial reasons, over a job that is less ‘degrading’.¹²¹ Moreover, we know that plenty of women find great satisfaction and fulfilment in temporarily using their bodies instrumentally. As one surrogate mother has expressed: ‘For us, giving someone a baby is as noble as giving a kidney to someone who needs it.’¹²²

In addition, it is equally plausible that surrogacy may help in deconstructing the harmful preconception that a woman is expected to stay in the home to care for a child simply because she gave birth to it. Further, by being paid for that labour, women are in fact being appropriately paid for what would otherwise be

¹²⁰ Radin (n 1) 1930.

¹²¹ *ibid.*

¹²² George Annas, ‘Death Without Dignity for Commercial Surrogacy: The Case of Baby M’ (1998) 18(2) HCR 21, 23.

voluntary work, which is uncontroversial in other instances where one sacrifices their time in the interest of another. It is not for anyone else to assert how women *should* use their body, which would in and of itself be contrary to the principle of bodily integrity which, by definition, incorporates the recognition of a woman's autonomy and self-determination over their own body.¹²³

Importantly, there is already some degree of inequity in the exchange between organ donor and donee, and surrogate and commissioning parents. First, as Erin and Harris observe, '[t]here is a lot of hypocrisy about the ethics of buying and selling organs and indeed other body products and services... What it usually means is everyone is paid but the donor.'¹²⁴ Indeed, the surgeons, wider medical team and transplant coordinator are remunerated, and the recipient receives an important benefit in kind. Only the donor is supposed to put up with the insult of no reward, in addition to shouldering the injury of the operation.

Finally, *because* the objective advantage of one party significantly outweighs the other, the people who are in dire need of the organs or who desire a child are *wholly* dependent upon the benevolence of another person. Permitting an exchange where both parties have something to objectively gain would both incentivise more people to volunteer, as well as relieve the feelings of guilt often experienced by the recipients of the gesture where they are unable to give anything in return.

¹²³ 'Bodily Integrity' (*Child Rights International Network*) <<https://archive.crin.org/en/home/what-we-do/policy/bodily-integrity.html>> accessed 10 March 2021.

¹²⁴ Erin and Harris (2003) (n 107) 137.

III. The Significance of Altruism

A final argument against commercial surrogacy or organ selling is that the marketisation of such acts would take away from their symbolic value as gifts that we can give each other in order to reinforce our sense of community.¹²⁵

However, in using the word ‘gift’, even sceptics such as Murray implicitly recognise that there is always some objectification implied when we offer our bodies or parts of our bodies to be used for the benefit of another.¹²⁶ To allow the option to either sell or donate would not necessarily take away from the value of the gift where the individual has opted to donate it. Sellers would know they have saved a life and would be appropriately compensated for their risk, time and altruism. This would not be diminished by sale. After all, we do not regard medicine as any the less a caring profession because doctors are paid. Rather, it may incentivise a greater number of organs being made available.

Thus, the argument that commercialising surrogacy and organ donation would detract from their inherent value as a means of fostering interpersonal relationships fails to consider that there is room and need for both. Those instances where a donor wants to be altruistic, and provide an organ or their womb completely free, will continue to exist. However, to prohibit a regulated market on this basis overlooks the deaths that occur as a consequence of the shortage of available organs and the expenses that prospective parents incur by travelling abroad to find surrogacy agencies. *Clearly*, this ideal of fostering solidarity through the vehicle of altruistic surrogacy and organ donation is

¹²⁵ Murray (1987) (n 114).

¹²⁶ *ibid.*

not powerful enough to persuade enough individuals to shoulder that burden themselves.

5. *Proposal for Reform*

A. Defending the Regulatory Approach to Saviour Siblings

Governments throughout the world have taken different approaches to regulating preimplantation genetic diagnosis (PGD). PGD enables parents with a known genetic abnormality to perform tests on an embryo in order to determine whether it carries a genetic abnormality. On the one hand, Ireland and Italy prohibit using PGD for saviour sibling selection.¹²⁷ On the other hand, in the US, there is a virtually free market for PGD subject

¹²⁷ In Ireland, Article 40.3.3 of the Irish Constitution expressly protects the right to life of the unborn. In Italy, the law prohibited selecting the embryos to be implanted, only allowing genetic counselling to couples if 'severe and irreversible abnormalities' are detected (Article 13 of law 40/2004). However, there are some recent court and tribunal rulings in Italy that have held that couples have a constitutional right to PGD to screen for genetic disease. See Emanuela Turillazzi and Vittorio Fineschi, 'Preimplantation Genetic Diagnosis: A Step-by-Step Guide to Recent Italian Ethical and Legislative Troubles' (2008) 34 JME 21, 1. Further, the European Court of Human Rights ruled that an Italian couple had an international human right access to PGD to screen for cystic fibrosis: *Costa v Italy* (European Court of Human Rights, chambers, Application No 54270/10, 28 August 2012).

to professional regulation and general criminal and civil law.¹²⁸ The UK falls between these extremes.

The HFEA Code of Practice requires clinics to consider the circumstances of each application for saviour siblings independently.¹²⁹ Both the consequences for any child born as a result of PGD and the family circumstances must be considered.¹³⁰ Such provisions ensure a minimum level of protection to not only the saviour sibling, but to the family unit as a whole. Furthermore, the policy requires that this detailed information be passed onto the family. It even requires that clinics put arrangements in place to enable long-term medical and psychosocial follow-up studies of children born as a result.¹³¹ Finally, as of January 2010, the HFEA upheld its case-by-case approach to licensing saviour siblings.¹³²

Importantly, the regulatory approach to saviour siblings consists of both hard law (legislation) and soft law (policy guidelines). Legislation is the most effective way of avoiding inconsistent decision-making by the HFEA.¹³³ Further, as

¹²⁸ There is no federal regulation of ART in the US and there are no laws explicitly addressing saviour sibling selection.

¹²⁹ HFEA, *Code of Practice* (n 105) [10.18].

¹³⁰ *ibid* [10.19]-[10.21].

¹³¹ *ibid* [10.19]-[10.24].

¹³² HFEA, *Minutes of Authority Meeting* (20 January 2010) [10:13].

¹³³ For example, see the different decisions reached in the Hashmi and Whitaker cases. The former case, also known as *Quintavalle*, is discussed in detail in section 1. In the latter case, the HFEA refused a licence for the Whitaker family to use PGD to solely conceive a direct tissue match for their existing child, Charlie, who had a non-inherited strain of Diamond-Blackfan anaemia. As the Whitakers were not at risk of passing on a heritable disease, there was no reason to screen for disease in conjunction with tissue typing. The reasoning was that the invasive and potentially harmful procedure could not be justified where the embryo did not also benefit in the sense of being free from a

highlighted by Brownsword, judicial determinations on new assisted reproductive technologies raise problems of incrementalism.¹³⁴ Indeed, we can find evidence of this in the cases concerning what constitutes ‘reasonable expenses’ in the surrogacy cases. The benefits of using legislation are that the broad guiding principles established and the boundaries regarding public interest concerns are subject to the rigours of parliamentary debate. Hard law then codifies these broad principles and ensures that there are formal monitoring mechanisms. Thus, the current regulatory approach rightly sanctions controversial applications of PGD via legislation.

Further, soft law, in the form of detailed guidance, provides for the relevant considerations to which clinics should give weight. As noted in section 1, the HFEA already establishes a multi-factorial approach which looks not only to the welfare of the saviour sibling, but to that of the family as a whole. The benefits of soft law are that, within the parameters set out by hard law, there is some degree of flexibility. This permits the clinics to ‘evaluate each contested commodification in its temporal and social context’.¹³⁵ These guidelines can evolve more easily than legislation, thereby resulting in detailed, practical and responsive information.

One criticism that has been made is regarding the HFEA’s lack of expertise, given its lay membership.¹³⁶ A solution

genetic disorder. This was a deeply flawed approach to take, as subsequently realised, as genetically screening an embryo for PGD does not directly confer any benefits on the embryo as that embryo has not been cured in any way. In both cases, the embryo is selected on the basis that it is a direct tissue match with an ill sibling.

¹³⁴ Roger Brownsword, ‘Reproductive Opportunities and Regulatory Challenges’ (2004) 67 MLR 304, 319-320.

¹³⁵ Radin (n 1) 1937.

¹³⁶ STC Report [196]-[198].

to this is to confer policymaking powers on a specialist advisory committee and to remove political control over the appointment of these individuals. An existing template that could improve the UK's regulatory approach is the Australian Health Ethics Committee (AHEC), which is a committee of the NHMRC. Its members have expertise in philosophy, law, clinical medical practice, health consumer issues, nursing, and disability.¹³⁷ A better approach would be to adopt a hybrid model, consisting of lay members as well as members of particular expertise. This is because specialists on the committee are likely to have different views on the subject. Further, it would allow for diversity of thought, thereby producing better outcomes. A hybrid model would also enhance the legitimacy of the committee, circumventing possible criticisms that the panel is elitist, and therefore illegitimate. Crucially, the policies of the committee will impact the lives of all, specialists and non-specialists. Therefore, it best serves the longevity of the committee for it to not only produce tailored policies, but to be seen to do so in a legitimate way.

B. Legalising Commercial Organ Donation and Commercial Surrogacy

A similar approach to that taken to regulating saviour siblings should be taken with commercial organ donation and surrogacy. With respect to organ donation, an ethical market would be confined to a self-governing geopolitical area. This could entail either a country, such as the UK's NHS, or the European Union. The UK's departure from the EU should not affect this as,

¹³⁷ NHMRC, *Australian Health Ethics Committee* (NHMRC 2012) <<https://www.nhmrc.gov.au/about-us/leadership-and-governance/committees/australian-health-ethics-committee-ahec>> accessed 23 January 2021.

notwithstanding this departure, the UK will continue to have a role in four other EU programmes.¹³⁸ Those who reside in that territory could sell into the system and they and their families would be eligible to receive organs. Much like the current system of the NHS, organ vendors would be contributing to a system which would benefit them and their families because their chances of receiving an organ in the case of need would be greatly increased.

One way to mitigate the risk of exploitation of low-income countries and their populations is to prohibit direct sales or purchasers. Only licensed purchasers, such as the NHS, would be able to buy these organs. Intermediary for-profit organisations such as agencies would not be permitted. Such intermediaries would merely prevent the donor from obtaining the full remuneration they would otherwise receive and would require the licensed bodies, such as the NHS, to pay more than they otherwise would have done. Moreover, it would dilute the accountability of those licensed purchasers to ensure that the vendor is not being exploited. By limiting who could purchase the organs, it would be easier to ensure that the bodily material could be adequately tested for harmful agents, such as HIV. Further, their provenance would be known. Such a system may also require sanctions to prevent abuse. The parameters set out above would be most effectively implemented via hard law.

Guidelines could be instituted in order to specify the kinds of considerations that licensed purchasers *must* follow. This would include similar provisions to those found in the HFEA Code of Practice, which requires clinics to consider the

¹³⁸ Namely, the nuclear research programme, the ITER to build the world's first functioning nuclear fusion system, the earth monitoring project Copernicus, and the EU satellite surveillance and tracking services.

circumstances of each application for saviour siblings independently.¹³⁹ This would include the consideration of the effect that the sale would have on the welfare of the donor¹⁴⁰ and on the donor's family, as found in the HFEA.¹⁴¹

Similarly, it would be possible to establish licensed surrogacy agencies, such as those found in the US. Although the courts in the UK already condone payment which surpasses reasonable expenses,¹⁴² they have done so by distorting the meaning of s 54(8) HFE Act 2008. This is because the presumption that payment which exceeds reasonable expenses is at odds with the best interests of the parties frequently does not withstand logical scrutiny. Legalising commercial surrogacy, when done via licensed agencies, would mitigate the risk of exploitation of low-income countries and their populations because prospective parents would not need to travel to such countries.

Moreover, hard law should be used to create implied duties that can be relied upon by the surrogate, should they be treated as a mere commodity. Such a duty would require the commissioning parents to actively take part in the general well-being of the surrogate before, during, and after giving birth to the child. Crucially, these should not be waived even if desired by the surrogate, much like the impermissibility of waiving certain terms in the Consumer Rights Act 2015 in contract law or non-delegable duties in tort. Such an approach would mitigate the risk inherent in commercial surrogacy of the surrogate being treated as of merely instrumental value.

¹³⁹ HFEA, *Code of Practice* (n 105) [10.18].

¹⁴⁰ HFEA Code 2018 (n 8) [10.25].

¹⁴¹ HFEA Code 2018 (n 10) [10.9].

¹⁴¹ *ibid.*

¹⁴² See section 1 for the full discussion of the relevant case law.

Conclusion

This article analysed the legal frameworks and ethical issues relating to the creation of saviour siblings, commercial organ donation and commercial surrogacy. After first setting out the law, the overarching ethical arguments against commodifying the body were considered. It was argued that these arguments rest on intuitions, grounded in dignity and the sacredness of life, that stifle rather than advance the discussion. Moreover, this article argued that the view that the separation of the body harms personhood rests on a conflation between the body as a whole as a vessel for personhood and the body as the sum of specific body parts. To insist that commodifying the body would do harm to personhood is incongruent with the biological reality of the actual importance of certain organs to the wellbeing of the donor and illiberally superimposes the view that such body parts are integral to personhood, thereby paternalistically undermining the autonomy of the donor. The argument that concern for personal autonomy does not entail anything like the extreme conception often associated with liberalism was accepted and held to be consistent with the qualified right to sell one's organs and surrogacy services argued for in this article. Furthermore, the risks and harms associated with commodification were evaluated with respects to the creation of saviour siblings, commercial organ donation and commercial surrogacy. It was concluded that the concerns regarding the commercialisation of organ donation and surrogacy are overstated and can be mitigated by establishing a regulatory framework similar to that which currently regulates the creation of saviour siblings.

