Autumn Vol. 72 No. 3 (2021)

NORTHERN IRELAND LEGAL QUARTERLY

NORTHERN IRELAND
LEGAL QUARTERLY

JOURNAL INFORMATION

The *Northern Ireland Legal Quarterly* is a leading peer-reviewed journal that provides an international forum for articles, commentaries and notes in all areas of legal scholarship and across a range of methodologies including doctrinal, theoretical and socio-legal. The journal regularly publishes special issues within this broad remit.

Established in 1936, the journal has a history and rich vein of legal scholarship, combining distinct publications on the law of Northern Ireland, and prominence within the School of Law at Queen's University Belfast, with leading contributions to the discussion and shaping of law across the common law world and further afield. The School of Law at Queen's University Belfast took over the publication of the journal from SLS Legal Publications (NI) Ltd in 2008, where it has since been published quarterly. The journal became an online-only publication in January 2017.

ISSN 2514-4936 (online) 0029-3105 (print)

AVAILABILITY AND ARCHIVES

The *Northern Ireland Legal Quarterly* is committed to making its contents widely available, to broaden our readership base. At least one article per issue is made available on an open access basis and may be published in advance. All articles become available on an open access basis on our website one year after publication.

All contributions to the journal become available on HeinOnline one year after publication (with issues going back to its launch in 1936) and LexisNexis three months after publication (with issues from 2019). The journal's contents appears on a growing range of indexing and abstracting services.

Since 2018 the journal's contents is promoted via social media and the Contributors' Blog.

In the summer of 2020, we expanded the reach and use of the *Northern Ireland Legal Quarterly* by adding 17 more years of content to the journal's existing archives. These now go back to 1999 (volume 50) and are widely accessed by our readership. Visit our Archive pages for further details.

SUBMISSIONS

The journal welcomes submissions of articles, commentaries, notes and book reviews on a rolling basis. Please see our 'For Authors' section for further details.

If you have any queries about the suitability of your article for the journal or if you have an idea for a special issue, please contact the Chief Editor Dr Mark Flear. For the contribution of commentaries and notes, please contact Dr David Capper. For book reviews, contact Dr Clayton Ó Néill.

SUBSCRIPTIONS

Subscriptions pay for a minimum of three months of exclusive access to the journal's latest contents (and up to one year for those who do not have access to LexisNexis).

NORTHERN IRELAND
LEGAL QUARTERLY

Autumn Vol. 72 No. 3 (2021)

Contents

Articles

Commentaries and Notes

Northern Ireland
Legal Quarterly
Autumn Vol. 72 No. 3 (2021) 411–447
Article DOI: 10.53386/nilq.v72i3.881

The Pandemic Influenza Preparedness Framework as a 'specialized international access and benefit-sharing instrument' under the Nagoya Protocol

Michelle Rourke

Griffith University, Brisbane

Mark Eccleston-Turner

King's College London

Correspondence email: m.rourke@griffith.edu.au

ABSTRACT

The World Health Organization (WHO) is starting to come to terms with the public health implications of the United Nations Convention on Biological Diversity (CBD) and its supplementary Nagoya Protocol about genetic resource access and benefit-sharing (ABS). Since 2017 there have been calls to recognise the WHO's Pandemic Influenza Preparedness (PIP) Framework as a specialized international ABS instrument under the Nagoya Protocol. This article will examine whether the PIP Framework meets the criteria of a specialized international ABS instrument as laid out in a 2018 study commissioned by the Subsidiary Body on Implementation to the CBD. Our analysis concludes that, while the PIP Framework meets the specialization criteria, it fails to meet the supportiveness criteria and does not provide legal certainty for pandemic influenza virus ABS, and therefore cannot constitute a specialized instrument under the CBD. Furthermore, we demonstrate that recognition of the PIP Framework as a specialized instrument would not mean that the CBD and Nagoya Protocol no longer apply to influenza viruses with human pandemic potential as has been asserted, rendering the relationship between the three international agreements unclear. As the WHO grapples with how to regulate access to other (non-influenza) human pathogens and the fair and equitable sharing of benefits associated with their use, a full appreciation of what ABS means when applied to pathogens is essential.

Keywords: Nagoya Protocol; PIP Framework; specialized instrument; Convention on Biological Diversity; treaty interaction; international law.

INTRODUCTION

The Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization to the Convention on Biological Diversity (Nagoya Protocol) is a binding international agreement that aims to ensure that the benefits arising from the utilisation of genetic resources are shared in a fair and equitable way, by governing access to those genetic resources, and by ensuring that justice, environmental concerns and sustainable development are incorporated into the transfer of those genetic resources from host countries to resource users. Article 4 of the Nagoya Protocol seeks to clarify its relationship with existing and future international agreements, recognising that parties may want to adopt certain access and benefit-sharing (ABS) measures that are specific to specialized subsets of genetic resources.[1] Articles 4(2) and 4(4) of the Nagoya Protocol thus create a designation of 'specialized international access and benefit-sharing instrument', or 'specialized instrument'. Unfortunately, there is no further guidance as to the criteria that such an instrument would need to meet in order to be considered a specialized instrument for the purposes of article 4 of the Nagoya Protocol.

There have been calls for the World Health Organization's (WHO) Pandemic Influenza Preparedness Framework for the Sharing of Influenza Viruses and Access to Vaccines and Other Benefits (PIP Framework) to be recognised as a specialized international ABS instrument under article 4 of the Nagoya Protocol,[2] but it is not yet clear whether the PIP Framework could be considered to be consistent with, and supportive of, the objectives of the United Nations (UN) Convention on Biological Diversity (CBD) and Nagoya Protocol. This article seeks to answer this question. It begins with an explanation of the PIP Framework and how it operates.[3] It next examines whether the PIP Framework can be considered a specialized international ABS instrument under article 4 of the Nagoya Protocol in accordance with the criteria outlined in the 2018 'Study into Criteria to Identify a Specialized International Access and Benefit-Sharing Instrument, and

1 Articles 4(2) and 4(3), Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization to the Convention on Biological Diversity 2010, UNEP/CBD/COP/10/27.

2 WHO, Review of the Pandemic Influenza Preparedness Framework: Report by the Director-General, A70/17, 10 April 2017, 23, Annex, Recommendation 36; WHO, Implementation of the Nagoya Protocol and Pathogen Sharing: Public Health Implications, EB140/15, 8 January 2017, 23.

3 For an analysis on the PIP Framework's SMTAs, see Michelle Rourke, 'Access by design, benefits if convenient: a closer look at the Pandemic Influenza Preparedness Framework's Standard Material Transfer Agreements' (2019) 97(1) The Milbank Quarterly 91.

a Possible Process for its Recognition' commissioned by the CBD's Subsidiary Body on Implementation[4] – the closest the international community has to a set of rules for specialized instrument recognition.

In doing so, we argue that the PIP Framework fails to meet the criteria to be recognised as a specialized instrument under article 4 of the Nagoya Protocol, as it does not adequately support the objectives of the CBD. We conclude that recognition of the PIP Framework as a specialized instrument would not mean that the CBD and Nagoya Protocol no longer apply to all influenza viruses with human pandemic potential, rendering the relationship between the three international agreements unclear. Furthermore, as the WHO decides how to regulate access to other (non-influenza) human pathogens and the fair and equitable sharing of benefits associated with their use, notably through a possible Pandemic Treaty, we intend for this article to contribute to a fuller understanding of the PIP Framework's ABS elements and the lessons that can be drawn about the use of the ABS mechanism in public health more generally.

SPECIALIZED INTERNATIONAL ABS INSTRUMENTS AND THE NAGOYA PROTOCOL

The Nagoya Protocol is a supplementary agreement under the UN CBD,[5] building on article 15 of the CBD that affirmed that countries have sovereignty over their genetic resources and have the authority to implement national legislation regulating their access and

4 CBD, Study into Criteria to Identify a Specialized International Access and Benefit-Sharing Instrument, and a Possible Process for its Recognition: Note By The Executive Secretary, CBD/SBI/2/INF/17, 29 May 2018; note that in 2013 Marie Wilke published two chapters that posed similar questions – whether the PIP Framework could be considered a specialized instrument under the Nagoya Protocol and whether it met the criteria of effectiveness and fairness. Her analyses will be referred to throughout. See Marie Wilke, 'A healthy look at the Nagoya Protocol – implications for global health governance' in Elisa Morgera, Matthias Buck and Elsa Tsioumani (eds), *The 2010 Nagoya Protocol on Access and Benefit-sharing in Perspective* (Martinus Nijhoff 2013); Marie Wilke, 'The World Health Organization's Pandemic Preparedness Framework as a public health resources pool' in Evanson Chege Kamau and Gert Winter (eds), *Common Pools of Genetic Resources – Equity and Innovation in International Biodiversity Law* (Routledge 2013). Since 2013 the discussions about what constitutes a specialized instrument have progressed within the CBD forum and the most recent study commissioned by the CBD starts to put a comprehensive structure around what does and does not constitute a specialized instrument. This analysis extends and updates Wilke's by using the newly stated criteria.

5 The CBD was opened for signature on 5 June 1992 and entered into force on 29 December 1993. The Nagoya Protocol to the CBD was adopted on 29 October 2010 and entered into force on 12 October 2014.

use.[6] The effect of the Nagoya Protocol was to elaborate on some of the uncertain definitions (such as 'derivatives' and 'utilisation'), develop some of the concepts (such as traditional knowledge associated with genetic resources) and introduce some of the machinery provisions for implementing the CBD (such as checkpoints, certificates of origin and so on). There are currently 193 contracting parties to the CBD and 123 parties to both the CBD and Nagoya Protocol,[7] essentially establishing three schemes: the CBD alone (73 Contracting Parties); the CBD plus Nagoya Protocol (123 Parties); and neither the CBD nor the Nagoya Protocol (United States and Holy See). As a minimum under the CBD and Nagoya Protocol, obligations include obtaining prior informed consent[8] and coming to mutually agreed terms about the use of genetic resources,[9] which can include the sharing of monetary or non-monetary benefits such as technology transfer, training and intellectual property.[10] This exchange of access to sovereign genetic resources in return for benefits associated with their use is known as access and benefit-sharing (ABS). While the Nagoya Protocol recognises that ABS can be achieved through multilateral mechanisms,[11] both the CBD and Nagoya Protocol envisage a bilateral contractual agreement between providers and users of genetic resources as the default ABS mechanism.

Article 4(2) provides:

> Nothing in this Protocol shall prevent the Parties from developing and implementing other relevant international agreements, including other specialized access and benefit-sharing agreements, provided that they are supportive of and do not run counter to the objectives of the Convention and this Protocol.[12]

Article 4(4) of the Nagoya Protocol continues:

> Where a specialized international access and benefit-sharing instrument applies that is consistent with, and does not run counter to the objectives of the Convention and this Protocol, this Protocol does not apply for the Party or Parties to the specialized instrument in respect of the specific genetic resource covered by and for the purpose of the specialized instrument.[13]

Inherent within article 4 is the acknowledgment that designation as

6 Art 15(1), Convention on Biological Diversity (Rio de Janeiro, 5 June 1992) 1760 UNTS 79, *entered into force* 29 December 1993.
7 As at 13 April 2020. See List of Parties.
8 Art 15(5).
9 Art 15(4).
10 Art 15(7).
11 Art 10.
12 Art 4(2).
13 Art 4(4).

a specialized instrument is a limited category. The Nagoya Protocol applies as the default ABS mechanism, and a specialized instrument can only deviate from the CBD/Nagoya Protocol ABS regime providing the instrument is 'supportive of and do[es] not run counter to the objectives of the Convention and this Protocol'.[14] Recognition that an agreement is a specialized international ABS instrument is significant in international law, not only because of the vague and undefined processes by which this is to occur, but because such recognition significantly alters the scope and operation of an already in-force treaty. If an agreement is classed as a specialized instrument under article 4 of the Nagoya Protocol, then the Nagoya Protocol no longer applies to the genetic resources included under that instrument to the extent that specialized instrument is consistent with the Nagoya Protocol.

There are currently two international instruments that might be considered specialized international ABS instruments under article 4 of the Nagoya Protocol.[15] The first is the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA or Plant Treaty) adopted by the UN's Food and Agriculture Organization (FAO) in 2001. The Plant Treaty was specifically developed 'in harmony' with the CBD[16] and governs ABS for a specific subset of agriculturally important plants.[17] The second is the WHO's PIP Framework. The PIP Framework was adopted as a World Health Assembly (WHA) Resolution in May 2011,[18] and outlines ABS arrangements for the subset of influenza viruses that have human pandemic potential,[19] but not for seasonal influenza viruses.[20] In this sense, the PIP Framework certainly appears to be a (highly) specialized ABS instrument. However,

14 Art 4(2).

15 Note also that negotiations for an international legally binding instrument (ILBI) under the UN's Convention on the Law of the Sea (UNCLOS) are ongoing and have included ABS measures for the marine genetic resources on the high seas, which may mean that the ILBI could become a specialized ABS instrument of sorts. However, it should be noted that genetic resources found on the high seas are outside the scope of the CBD and Nagoya Protocol, and therefore the ILBI would not require (nor presumably qualify for) recognition as a specialized instrument because the Nagoya Protocol only applies to sovereign genetic resources.

16 Art 1(1), International Treaty on Plant Genetic Resources for Food and Agriculture 2001, 2400 UNTS 303

17 Annex 1, and any other materials included within the Multilateral System (arts 11(1) 11(5) and 15(1)). Some parties to the CBD have already recognised the Plant Treaty as a specialized instrument in their domestic legislation. See WHO EB140/15 (n 2 above) 24.

18 WHO, Pandemic Influenza Preparedness: Sharing of Influenza Viruses and Access to Vaccines and Other Benefits (PIP Framework) UN Doc A64/VR/10, 24 May 2011.

19 Art 3(1).

20 Art 3(2).

unlike the Plant Treaty, the PIP Framework was not developed to be in harmony with the CBD.[21] In fact, the PIP Framework does not mention the CBD within its text, despite article 15 of the CBD providing the legal foundation for parts of the PIP Framework.[22] Therefore the PIP Framework's potential recognition as a specialized instrument under article 4 of the Nagoya Protocol is all the more complex and significant, given that the PIP Framework could cast the Nagoya Protocol into abeyance in the field of pandemic influenza viruses.

THE GLOBAL SHARING OF INFLUENZA VIRUS SAMPLES

Informal influenza virus sample sharing

Influenza poses a significant risk to the human population. Effectively combating pandemic influenza requires an internationally coordinated response which includes testing, surveillance, the development of antiviral medication and strain-specific vaccines. This is an ongoing process for influenza, as seasonal strains change year to year (genetic drift) and can recombine to produce potentially pandemic strains (pandemic shift) where new influenza subtypes emerge.[23] As a result, the international scientific community has been sharing influenza viruses informally for decades, monitoring the changing genetic sequence of seasonal strains and hoping to detect a pandemic strain before it starts to take hold in the human population. The WHO coordinates the sharing of virus samples between this network of laboratories that has existed in some form since the 1950s.[24] The sharing of seasonal and potentially pandemic influenza virus samples between the laboratories of the Global Influenza Surveillance Network (GISN)[25] occurred on an informal basis until 2006 when attitudes to informal virus sharing started to shift.

---

21 The PIP Framework and the Nagoya Protocol negotiations overlapped, with the two processes influencing each other. For a detailed explanation of how these negotiation processes interacted, see Wilke, 'A healthy look at the Nagoya Protocol' (n 4 above).

22 See the section on 'The development of the PIP Framework', page 417 below.

23 Antole Krattiger et al, 'Intellectual property management strategies to accelerate the development and access of vaccines and diagnostics: case studies of pandemic influenza, malaria, and SARS' (2006) 2(2) Innovation Strategy Today 67.

24 WHO (n 18 above); as Influenza A viruses infect multiple animal hosts, there is a similar network of laboratories that share animal influenza viruses coordinated by the FAO.

25 Renamed the Global Influenza Surveillance and Response System in 2011.

The development of the PIP Framework

In 2006, in response to the threat posed by H5N1 avian influenza virus, the WHA passed resolution 59.2, which called upon WHO member states to '[d]isseminate to the WHO collaborating centres information and relevant biological materials related to highly pathogenic avian influenza and other novel influenza strains in a timely and consistent manner'.[26] At the time, Indonesia had the highest number of infections and deaths from H5N1.[27] Despite this, and the established norm of free virus sharing between laboratories of the GISN, Indonesia's sharing of virus samples with GISN fluctuated between openly sharing samples and refusing to share, claiming that Indonesia had sovereign authority over the samples isolated within its territories, and that it was therefore under no obligation to share them with the wider international community.[28] In claiming that the virus samples were its sovereign resources, Indonesia invoked the CBD, which states that 'access to genetic resources shall be subject to prior informed consent of the Contracting Party providing such resources',[29] and any access granted 'shall be on mutually agreed terms'.[30]

This framing helped Indonesia highlight the inequity of being expected to share virus samples with GISN but not being afforded fair access to the vaccines and antivirals developed using those samples,[31] challenging the notion that the existing system of pandemic influenza preparedness was a global public good.[32] Indonesia's then Health Minister Siti Fadilah Supari claimed that the WHO transferred the samples Indonesia provided on to pharmaceutical companies to develop pandemic influenza vaccines, who then patented the vaccine and its components which developing countries could not afford.[33] The basis of this claim was subsequently shown to be correct when it

26 WHO, Application of the International Health Regulations (2005) WHA59.2, 26 May 2006.

27 WHO, Table A 2003–2009, Cumulative number of confirmed human cases for avian influenza A(H5N1) reported to WHO, 2003–2015 (2015).

28 Peter Gelling, 'Indonesia defiant on refusal to share bird flu samples' *New York Times* (New York, 26 March 2015) 12; Endang R Sedyaningsih et al, 'Towards mutual trust, transparency and equity in virus sharing mechanism: the avian influenza case of Indonesia' (2008) 37(6) Annals of the Academy of Medicine, Singapore 482.

29 Art 15(5).

30 Art 15(4).

31 Rachel Irwin, 'Indonesia, H5N1, and global health diplomacy' (2010) 3 Global Health Governance.

32 Jeremy Youde, *Globalisation and Health* (Rowman & Littlefield 2019) 115.

33 Declan Butler, 'Q&A: Siti Fadilah Supari' (2007) 450(7173) Nature 1137; Sedyaningsih et al (n 28 above).

was discovered that virus samples were being transferred from GISN laboratories to pharmaceutical manufacturers without consultation with, or the permission of, the originating country.[34] This controversial action forced the WHO to put formal terms around the GISN's virus-sharing practices in what was to become known as the PIP Framework.[35] During the negotiations, Indonesia submitted:

> [A] framework of benefit sharing is to be developed through agreed terms and conditions to ensure a global stockpile of pre-pandemic and pandemic vaccines, accessibility of vaccine at an affordable price, access to and transfer of technology and know-how for production of vaccines and empowerment and capacity building of vaccine manufacturing in developing countries.[36]

After four years of negotiations, the PIP Framework was adopted in May 2011.[37] It provides for recommendations in two areas: the timely sharing of influenza samples with human pandemic potential between member states and the WHO via the newly renamed Global Influenza Surveillance and Response System (GISRS);[38] and the sharing of virus samples with third-party entities that operate outside of the GISRS, such as pharmaceutical and vaccine manufacturers, in return for these external entities sharing benefits with the WHO for distribution to member states in the event of an influenza pandemic.[39] Thus, the PIP Framework is ostensibly an ABS framework governing access to viral genetic resources in exchange for the benefits arising from their use.

The Nagoya Protocol to the Convention on Biological Diversity

The CBD was adopted in 1992 and used by Indonesia in 2006 and 2007 as the legal basis for claiming sovereignty over influenza virus samples. Article 15.1 of the CBD 'recogniz[es] the sovereign rights of States over

34 Colin McInnes and Kelly Lee, *Global Health and International Relations* (Polity Press 2012) 193.

35 Jeanette Lange, 'Negotiating issues related to pandemic influenza preparedness: the sharing of influenza viruses and access to vaccines and other benefits' in Ellen Rosskam and Ilinoa Kickbusch (eds), *Negotiating and Navigating Global Health: Case Studies in Global Health Diplomacy* (World Scientific 2012).

36 WHO, Submission of the Government of Indonesia, 'Fundamental Principles and Elements for the Development of a New System for Virus Access and Fair and Equitable Benefit Sharing Arising from the Use of the Virus for the Pandemic Influenza Preparedness' at Interdisciplinary Working Group, 'Sharing of Influenza' A/PIP/IGM/WG/4, 3 April 2008.

37 64th World Health Assembly, Pandemic Influenza Preparedness: Sharing of Influenza Viruses and Access to Vaccines and Other Benefits (PIP Framework), Geneva, Switzerland: WHO (2011: WHA64.5).

38 Art 5(1).

39 Art 6(11).

their natural resources' and affirms that national governments have 'the authority to determine access to genetic resources'.[40] The CBD has three objectives:

> [1] the conservation of biological diversity, [2] the sustainable use of its components and [3] the fair and equitable sharing of the benefits arising out of the utilization of genetic resources, including by appropriate access to genetic resources and by appropriate transfer of relevant technologies, taking into account all rights over those resources and to technologies, and by appropriate funding.[41]

Indonesia argued that this third objective about accessing genetic resources and sharing the benefits associated with their use (abbreviated as 'access and benefit-sharing', or ABS), applied to viruses that were isolated from within their territorial borders.[42] Developed countries held that viruses were a threat to biodiversity and not the sort of genetic resource that ought to be regulated by an environmental conservation treaty.[43] By adopting the PIP Framework, which 'recognize[d] the sovereign right of States over their biological resources'[44] the WHO had implicitly accepted the premise of Indonesia's argument:[45] viruses are the sovereign genetic resources of nation states and are therefore subject to benefit-sharing obligations under the CBD.[46] The CBD clarifies that this means nation states have the authority to implement domestic measures regulating their genetic resources. Article 3 provides, in part: 'States have, in accordance with the Charter of the UN and the principles of international law, the sovereign right to exploit their own resources pursuant to their own environmental policies.'[47]

The negotiations for the PIP Framework at the WHO coincided with negotiations of the Nagoya Protocol at the CBD, however, the text of the PIP Framework remained silent regarding its relationship to both

40 Art 15(1).

41 Art 1.

42 Sedyaningsih et al (n 28 above).

43 See explanations in David Fidler, 'Influenza virus samples, international law, and global health diplomacy' (2008) 14(1) Emerging Infectious Diseases 88, 90; Frederick M Abbott, 'An international legal framework for the sharing of pathogens: issues and challenges' (2010) ICTSD Issue Paper No 30, 13–14; Frank L Smith III, 'Insights into surveillance from the influenza virus and benefit sharing controversy' (2012) 24(1) Global Change, Peace and Security 79.

44 Art 1(11).

45 Michelle Rourke, 'Restricting access to pathogen samples and epidemiological data: a not-so-brief history of "viral sovereignty" and the mark it left on the world' in Mark Eccleston-Turner and Ian Brassington (eds), *Infectious Diseases in the New Millennium: Legal and Ethical Challenges* (Springer 2020).

46 Michelle Rourke, 'Viruses for sale – all viruses are subject to access and benefit sharing obligations under the Convention on Biological Diversity' (2017) 39(2) European Intellectual Property Review 79–89.

47 Art 3.

the CBD and the Nagoya Protocol. This was despite the instruments regulating overlapping subject matter: human pandemic influenza viruses fall under the PIP Framework and are still within the remit of the CBD and Nagoya Protocol. Nevertheless, on the face of it, the PIP Framework does appear to be a specialized ABS instrument because it places rules and obligations around a specific set of genetic resources that also fall within the remit of the CBD and Nagoya Protocol: accessing influenza viruses with pandemic potential and sharing the benefits associated with their use.

The WHO's 2016 internal review of the PIP Framework specifically addressed the interaction of the PIP Framework with the Nagoya Protocol. It stated:

> The PIP Framework is a multilateral access and benefit sharing instrument that appears to be consistent with the objectives of the Nagoya Protocol … the implementation of the Nagoya Protocol may introduce uncertainty in relation to the sharing of influenza viruses, since numerous bilateral transactions could be required to be negotiated, which could delay the access to viruses. As more countries put in place domestic legislation to implement the Nagoya Protocol, the urgency increases to resolve this uncertainty and reduce the risk to global health security.[48]

The concern here is that countries may exercise their sovereignty by choosing to regulate ABS for human pandemic influenza viruses through national measures implementing the binding CBD or CBD/Nagoya Protocol agreements, rather than participating in the WHO's non-binding virus-sharing arrangements under the PIP Framework. In light of this uncertainty, the PIP Review Group recommended that '[t]he PIP Framework should be considered as a specialized international instrument to clarify the implementation of the Nagoya Protocol in relation to pandemic influenza preparedness and response'.[49] After considering the PIP Review Group's report, the seventieth WHA in 2017 adopted Decision WHA70(10) which 'reaffirme[d] the importance of the PIP Framework' and 'emphasize[d] its critical function as a specialized international instrument' for accessing pandemic influenza viruses and sharing vaccines and other benefits.[50] That is, it may function as a specialized ABS instrument at times, but it does not have any legal recognition as such.

48 WHO (n 18 above) annex, 22.
49 Ibid 23.
50 Ibid para 2.

THE INTERACTION OF THE PIP FRAMEWORK AND NAGOYA PROTOCOL

The parallel operation of the PIP Framework and Nagoya Protocol does create some confusion. To reiterate, article 3.1 of the PIP Framework states that it 'applies to the sharing of H5N1 and other influenza viruses with human pandemic potential and the sharing of benefits',[51] but functionally, the PIP Framework only applies to those 'H5N1 and other influenza viruses with human pandemic potential' that countries choose to share through the GISRS and thus become 'PIP biological Materials'.[52] Other samples of H5N1 and other influenza viruses with human pandemic potential exist outside of the GISRS network. They were never part of the network and remain outside of the regulatory reach of the PIP Framework. These virus samples are instead captured under the regulation of 'genetic resources' under article 15 of the CBD and the Nagoya Protocol.[53] Because the PIP Framework does not have official recognition as a specialized instrument under article 4(4) of the Nagoya Protocol, the viruses within the scope of the PIP Framework are also covered by the CBD and Nagoya Protocol.[54] This means that countries that have shared, or would usually share, pandemic influenza viruses through the PIP Framework, might instead choose to enter into bilateral arrangements with a vaccine manufacturer (or other party), removing the WHO as the intermediary, as per their rights under the CBD and Nagoya Protocol. Therefore, the parallel functioning of the PIP Framework and Nagoya Protocol could undermine the whole point of the PIP Framework and the functioning of the GISRS. Official recognition of the PIP Framework as a specialized instrument under article 4(4) of the Nagoya Protocol would theoretically mean that the CBD and Nagoya Protocol no longer apply to those viruses already within the scope of the PIP Framework. Hence the importance of recognition as a specialized instrument and the importance of working out precisely what a specialized instrument should be in order to have it qualify for such recognition.

It certainly appears that the PIP Framework is a specialized international ABS instrument insofar as the international community is using it as such. Between 1 December 2012 and 30 June 2019, 1205 PIP biological materials were recorded in the PIP Framework's

51 Art 3(1).

52 Arts 4(1) and 5(1). Note that not all viruses shared through the GISRS are 'PIP biological materials' as non-pandemic influenza and other viruses are also shared with this network of laboratories.

53 Arts 2 and 15, CBD; Art 3, Nagoya Protocol.

54 Art 4(4).

Influenza Virus Tracing Mechanism (IVTM),[55] although the number of influenza samples that pass through the GISRS each year far surpasses this number. However, recent problems with accessing influenza viruses through the GISRS indicate that the system is being sidelined in favour of alternative ABS arrangements.[56] The following analysis of the PIP Framework in light of article 4(4) of the Nagoya Protocol may point to some of the reasons why some users are avoiding it.

CRITERIA FOR SPECIALIZED INTERNATIONAL ABS INSTRUMENTS

In December 2016, the second meeting of the Conference of the Parties to the CBD Serving as the Meeting of the Parties to the Nagoya Protocol (COP-MOP 2) adopted Decision 2/5 on 'Cooperation with other international organizations, conventions and initiatives'.[57] During the discussions, 'some express[ed] concern over the initiative taken [by the WHO] outside of the [Nagoya] Protocol to clarify its relationship with the PIP Framework'.[58] Decision 2/5 requested the Executive Secretary of the CBD to 'conduct a study into criteria that could be used to identify what constitutes a specialized international access and benefit-sharing instrument' under article 4(4) of the Nagoya Protocol and investigate 'what could be a possible process for recognizing such an instrument'.[59] The resulting study[60] was presented via the Subsidiary

55 WHO, Pandemic Influenza Preparedness Framework Progress Report 1 January 2018–30 June 2019 (2019) 7.

56 WHO, Approaches to Seasonal Influenza and Genetic Sequence Data under the PIP Framework, 14 December 2018, 30; GISAID, GISAID's Comments on the WHO Report of the Public Health Implications of Implementation of the Nagoya Protocol, 13 May 2019.

57 CBD, 'Second Meeting of the Conference of the Parties to the Convention on Biological Diversity Serving as the Meeting of the Parties to the Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization: Decision 2/5 Adopted by the Parties to the Nagoya Protocol on Access and Benefit-Sharing: Cooperation with other international organization, conventions and initiatives' CBD/NP/MOP/DEC/2/5, 16 December 2016.

58 Stephanie Switzer et al, 'Biodiversity, pathogen sharing and international law' in Stefania Negri (ed), *Environmental Health in International and EU Law: Current Challenges and Legal Responses* (Routledge 2019) 281.

59 CBD (n 57 above) para 3, 2.

60 CBD (n 4 above).

Body on Implementation to the COP-MOP 3 in 2018.[61] The Study into the Criteria to Identify a Specialized International Access and Benefit-Sharing Instrument, and a Possible Process for its Recognition (hereafter 'the Study') outlined nine potential criteria for specialized international ABS instruments under two categories: specialization and supportiveness).[62] The specialization criteria refer to the extent to which an instrument addresses specific uses of genetic resources which would require a differentiated (from the Nagoya Protocol) and hence specialized approach, whereas supportiveness refers to the extent to which the instrument is consistent with the aims, objectives and approach of the CBD and Nagoya Protocol.

This section addresses the criteria laid out in the 2018 Study with specific reference to the objectives, text and operation of the PIP Framework and its associated Standard Material Transfer Agreements (SMTAs).[63] The 2016 Review of the PIP Framework stated that the PIP Framework 'appears to be consistent with the objectives of the Nagoya Protocol'.[64] This section analyses the extent to which the PIP Framework, as the world's only virus-specific ABS instrument, is actually consistent with the objectives of the Nagoya Protocol by examining each of the Study's criteria in turn, under the Study's categories of specialization and supportiveness. While the Study does not 'necessarily reflect the views of the [CBD] Secretariat'[65] and the criteria for specialized instruments are still very much under discussion,[66] it does provide a point of reference for further discussion. A thorough examination of the PIP Framework using these criteria is instructive for those wishing to improve the ABS process for those genetic resources requiring a specialized approach outside of the CBD and Nagoya Protocol's default bilateral contract arrangements.

61 CBD, 'Third Meeting of the Conference of the Parties to the Convention on Biological Diversity Serving as the Meeting of the Parties to the Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization: Decision 3/14 Specialized international access and benefit sharing instruments in the context of Article 4, paragraph 4, of the Nagoya Protocol' CBD/NP/MOP/DEC/3/14, 31 January 2019, para 1, 1.

62 CBD (n 4 above) 12–13.

63 Annex 1 (Standard Material Transfer Agreement 1) and annex 2 (Standard Material Transfer Agreement 2), PIP Framework (n 18 above).

64 WHO (n 18 above) annex, 22.

65 CBD (n 4 above) 1.

66 CBD (n 61 above) annex, 2.

Specialization

An instrument under consideration would be intergovernmentally agreed upon

The PIP Framework was adopted as Resolution WHA64.5 by the WHA in May 2011, in accordance with article 23 of the Constitution of the WHO, giving it the status of a 'Resolution of the World Health Assembly'.[67] Resolutions require a two-thirds majority of the 196 member states before passing. The PIP Framework is certainly intergovernmentally agreed upon 'under the aegis of an international organization created by a treaty'.[68] But this criterion highlights the perhaps obvious point that the membership of the WHO is not the same membership as that of the Nagoya Protocol, and the Meeting of the Parties to the Nagoya Protocol is likely the only body that could adopt or accept the PIP Framework as a specialized instrument.[69]

An 'instrument' may be either binding or non-binding

The PIP Framework takes the form of a resolution of the WHA, as opposed to a treaty. This is not inherently a limitation to the PIP Framework being recognised as a specialized instrument: the language of article 4(4) is particularly broad in referring to 'instruments' as opposed to 'agreements', implying that the states parties of the Nagoya Protocol anticipated a scenario whereby instruments not grounded in treaty law could constitute a specialized instrument. Thus, the PIP Framework meets the second criterion of the Study.

However, Wilke points out that the PIP Framework's non-binding nature is not irrelevant, as it has implications for the scope of the instrument and the background functioning of the CBD and Nagoya Protocol's ABS regime. Wilke states that the PIP Framework 'only functions as a specialized ABS instrument for influenza viruses where transfers are covered by the Framework's binding contract clauses',[70] meaning just the ones that countries have chosen to share with the GISRS through the mechanisms created by the PIP Framework. As a result, the CBD and Nagoya Protocol remain the default ABS mechanism regulating the transfer of pandemic influenza viruses that member

67 Constitution of the World Health Organization (New York, 22 July 1946) 14 UNTS 185, *entered into force* 7 April 1948.

68 CBD (n 4) 6.

69 This is the only agreement discussed here with a provision on specialized instruments (the CBD does not have any equivalent provision). Therefore, the acceptances of the PIP Framework as a specialized instrument would likely only impact parties to the Nagoya Protocol.

70 Wilke, 'A healthy look at the Nagoya Protocol' (n 4 above) 126.

states do *not* share with the GISRS through the PIP Framework (as they are free to do).

All this is to say that, if the PIP Framework were binding, it would be likely the default arrangement for the sharing of pandemic influenza viruses; there would be a clear international obligation on member states to share these viruses with the WHO under the PIP Framework. The only point at which parties take on the obligations of the PIP Framework's ABS provisions is when they enter into SMTAs with the WHO, and member states can choose which virus samples to share under the SMTAs.[71] This means that the PIP Framework does not actually create any obligation to share pandemic influenza viruses with the GISRS. Member states can continue to enter into bilateral ABS arrangements under the CBD and Nagoya Protocol, and, thus, no influenza viruses can be 'considered exempt from the Nagoya Protocol's scope by virtue of Article 4.4'.[72]

The implication of this is that, even if the PIP Framework is considered a specialized instrument under article 4(4), it does not follow that all influenza viruses with human pandemic potential automatically fall outside of the scope of the Nagoya Protocol – only the ones which member states actively choose to transfer to GISRS through the mechanisms created by the PIP Framework itself. The disapplication of the Nagoya Protocol for those virus samples would make very little sense, however, when the PIP Framework itself states that 'member states may also provide PIP biological materials directly to any other party or body on a bilateral basis provided that the same materials are provided on a priority basis to the WHO'.[73]

An instrument would apply to a specific set of genetic resources and/or traditional knowledge associated with genetic resources, which would otherwise fall under the scope of the Nagoya Protocol

The CBD defines 'genetic resources' as 'genetic material of actual or potential value'.[74] The term 'genetic material' is defined as 'any material of plant, animal, microbial or other origin containing functional units of heredity'.[75] The Nagoya Protocol uses the same definitions for these terms as provided for in article 2 of the CBD.[76] Viruses, as protein capsules containing DNA or RNA ('genetic material') which are useful in scientific research and the development of vaccines and other products

71 Ibid 145.
72 Ibid 146.
73 Art 5(1)(4).
74 Art 2, CBD
75 Ibid.
76 Art 2, Nagoya Protocol.

('of actual or potential value'), do fit within the CBD's (and therefore the Nagoya Protocol's) definition of 'genetic resources'.[77] If countries so choose, they can implement legislative, administrative and policy measures for accessing viral genetic resources, including provisions on prior informed consent and benefit-sharing as part of mutually agreed terms. The PIP Framework applies to a very narrow subset of the world's known viruses, only 'H5N1 and other influenza viruses with human pandemic potential'.[78] It does not apply to seasonal influenza viruses or other non-influenza pathogens.[79] The PIP Framework therefore applies to a specific set of genetic resources that would otherwise fall within the scope of the Nagoya Protocol, meeting the third criterion for a specialized instrument.

The third and fourth criteria from the Study address traditional knowledge (TK) associated with genetic resources, a uniquely challenging aspect of ABS under the CBD and Nagoya Protocol. The PIP Framework does not address the issue of TK associated with H5N1 and other influenza viruses with human pandemic potential. It may seem easy to discount the relevance of the TK of Indigenous Peoples and local communities associated with influenza viruses as it has not previously factored into the sharing of influenza virus samples through the WHO's GISRS network of laboratories. But there is an increasing understanding of the importance of TK associated with genetic resources, including viruses.[80]

The CBD asks contracting parties to 'respect, preserve and maintain knowledge, innovations and practices' of Indigenous Peoples and local communities, and 'encourage[s] the equitable sharing of the benefits arising from the utilization of such knowledge, innovations and practices'.[81] The Nagoya Protocol goes much further, creating similar obligations for the use of TK 'held or owned'[82] by Indigenous Peoples and local communities associated with genetic resources as for the

77 Rourke (n 45 above) 79–89. Note also that if viruses did not fit within the scope of the CBD and Nagoya Protocol, there would not be a push from the WHO to have the PIP Framework included as a specialized instrument within this ABS regime. This in itself is evidence that the international community has decided to treat viruses as being subject to the ABS provisions of the CBD and Nagoya Protocol.

78 Art 3(1).

79 Art 3(2).

80 See Michelle Rourke, 'When knowledge goes viral: assessing the possibility of virus-related traditional knowledge for access and benefit-sharing' (2018) 21(5–6) Journal of World Intellectual Property 356.

81 Art 8(J).

82 Preamble, Nagoya Protocol.

genetic resources themselves. This includes prior informed consent for accessing TK[83] and the sharing of benefits on mutually agreed terms.[84]

The definitions of both 'traditional knowledge' and 'Indigenous and local communities' in the CBD and Nagoya Protocol ABS regimes can be broadly interpreted[85] and, as with most of the contentious issues in this arena, have been left to individual parties to determine in their own legislative, administrative and policy measures. This means that there are many forms of TK that could be considered to be associated with viral genetic resources, including traditional burial practices that were associated with the spread of the Ebola virus during the West African Ebola epidemic in 2014–2015,[86] and in the context of influenza viruses 'it is possible that knowledge about traditional poultry or pig farming practices that provide insights about a strain of influenza virus (its transmission, virulence, or the susceptibility of particular animals) might qualify' as TK, the use of which may require prior informed consent and mutually agreed terms.[87]

Traditional medicine has always been within the purview of the WHO, although it did not address the intellectual property aspects of traditional medicines until December 2000.[88] The Inter-regional Workshop on Intellectual Property Rights in the Context of Traditional Medicine 'produced recommendations regarding, *inter alia* ... the equitable sharing of benefits for commercial use of traditional medicine'.[89] Since 2000, the WHO has been working with the World Intellectual Property Organization (WIPO) to address 'the need to prevent misappropriation of health-related traditional knowledge' in the context of traditional medicine.[90] Thus, the WHO had been engaged in these issues at least a decade before the adoption of the

83 Arts 6(2) and 7.

84 Arts 5(2) and 7. Such measures are, of course, softened with the vague language that pervades the Nagoya Protocol: 'In accordance with domestic law, each Party shall take measures, as appropriate ...'. It is *not* clear whether not taking measures could also be considered 'appropriate' under the Nagoya Protocol.

85 Michelle Rourke, 'Who are "indigenous and local communities" and what is "traditional knowledge" for virus access and benefit-sharing? A textual analysis of the Convention on Biological Diversity and its Nagoya Protocol' (2018) 25(3) Journal of Law and Medicine 707–726.

86 Angellar Manguvo and Benford Mafuvadze, 'The impact of traditional and religious practices on the spread of Ebola in West Africa: time for a strategic shift' (2015) 22 (Suppl 1):9 Pan African Medical Journal 1–4.

87 Rourke (n 85 above).

88 Silke von Lewinski, *Indigenous Heritage and Intellectual Property: Genetic Resources, Traditional Knowledge and Folklore* (2nd edn, Kluwer Law International 2008) 45.

89 Ibid.

90 Dr Margaret Chan's Address at the WHO Congress on Traditional Medicine (WHO 2008).

PIP Framework. Given the primacy of TK in the Nagoya Protocol and the WHO's engagement on the issue of benefit-sharing related to the use of TK since 2000, it might be considered a major shortcoming of the PIP Framework that it does not even address the *possibility* of TK associated with influenza viruses with human pandemic potential.

An instrument would apply to specific uses of genetic resources and/or traditional knowledge associated with genetic resources, which would require a differentiated and hence specialized approach

As stated, prior to the adoption of the PIP Framework in 2011, the sharing of influenza viruses between the international scientific community and coordinated by the WHO was informal, but still highly structured and systematised. The basic structure of the GISN/GISRS has remained the same since 2011,[91] but has since expanded and increased its global reach. The genetic properties of influenza viruses and the ongoing requirement to monitor their evolution (genetic drift) and the emergence of pandemic strains (through genetic shift) have dictated how the GISN/GISRS is structured. As a simplification, the GISRS laboratory network consists of 144 National Influenza Centers (NICs) which together 'process more than 3 million clinical specimens globally every year'[92] and continuously feed information and virus samples into regional Collaborating Centers (CCs).[93] These, in turn, deliver information to the WHO and provide candidate vaccine virus isolates to vaccine manufacturers. This is the only global influenza laboratory network of its kind, and its continued operation is essential to determining which seasonal strains should be used to make annual influenza vaccinations and to detect the emergence of potentially pandemic strains. It is clear that the sharing of influenza viruses requires a different, specialized approach to that of other genetic resources, and even to that of other pathogens. But whether or not the PIP Framework can be said to constitute that specialized approach is rather more complicated.

Prior to 2011, the transfer of clinical samples and virus isolates from the NICs to the CCs occurred on an informal basis, and vaccine candidate virus isolates that originated from NICs were provided to vaccine manufacturers from CCs free of charge. The PIP Framework

91 The change of name from GISN to GISRS was symbolic. Youde notes that it 'helped to convey the message that a change had actually occurred as part of the [PIP Framework] negotiations': Youde (n 32 above) 130.

92 WHO (n 56 above) 7.

93 Alan Hay and John McCauley, 'The WHO global influenza surveillance and response system (GISRS): a future perspective' (2018) 12(5) Influenza and Other Respiratory Viruses 551. The GISRS network also includes Essential Regulatory Laboratories (ERLs).

had to be retrofitted to this 'pre-existing monopoly', where access to viruses was already controlled and tracked through the WHO's global network of laboratories.[94] The PIP Framework created a formal administrative structure around the transfer of human clinical specimens, wild-type and modified influenza viruses with human pandemic potential (now called 'PIP biological materials')[95] between WHO-affiliated laboratories (NICs and CCs) using the SMTA1, and from WHO laboratories to third parties, like vaccine manufacturers, using the SMTA2. Importantly, though, this formal administrative ABS structure only applies to the subset of virus samples that are identified as having human pandemic potential. And so, we would question why this specialized ABS approach is required for pandemic influenza viruses but not the many more seasonal influenza viruses that are shared in the same manner, through the same GISRS, but without the application of the PIP Framework.

Supportiveness

The remaining five criteria for specialized instruments refer to the extent to which the instrument under consideration is mutually supportive of the aims, objectives and functioning of the CBD and Nagoya Protocol. It is worth noting that 'mutual supportiveness' is a somewhat underdeveloped concept and '[l]ittle is known ... as to the exact legal nature and scope of the concept and its role in treaty interpretation and adjudication at the international and national level'.[96] Despite this, Wilke has commented that 'the two instruments [the PIP Framework and Nagoya Protocol] seem to be well equipped to form a working symbiosis'.[97] The extent to which this apparent symbiosis is sufficient to meet the criteria for specialized instruments under the 2018 Study will be considered below.

Consistency with biodiversity conservation and sustainable use objectives

When Indonesia withheld its influenza virus samples from the WHO in 2006/2007, many commentators highlighted the incompatibility of applying the sovereignty provisions of the CBD, an environmental conservation treaty, with public health objectives which include the

94 Deborah Scott and Dominic Berry, 'Genetic resources in the age of the Nagoya Protocol and gene/genome synthesis' (Report and Analysis of an Interdisciplinary Workshop 18 November 2016) 20.

95 Art 4(1).

96 Wilke, 'A healthy look at the Nagoya Protocol' (n 4 above) 137.

97 Ibid 146.

eradication of disease agents like influenza viruses.[98] Fidler, for example, stated that 'interpreting the CBD to apply to pathogenic viruses may be contrary to the CBD's purpose',[99] highlighting (correctly) that 'avian influenza viruses ... are not the kind of biological and genetic resources that the CBD sought to protect and regulate through the principles of sovereignty, prior informed consent, and mutual benefits from access and exploitation'.[100] But arguments of this nature[101] tend to take into account only the first two objectives of the CBD: 'the conservation of biological diversity' and 'the sustainable use of its components'.[102] Indonesia's argument and the ensuing PIP Framework undoubtedly fit within the third objective of the CBD: 'the fair and equitable sharing of the benefits arising out of the utilization of genetic resources'.[103] It is this third objective around which the entire Nagoya Protocol is built.[104]

The text of the CBD does not *explicitly* link its third objective of 'fair and equitable sharing of the benefits arising out of the utilization of genetic resources' with achieving the first two objectives of biodiversity conservation and sustainable use. However, the Nagoya Protocol's objective is to achieve:

> ... fair and equitable sharing of the benefits arising from the utilisation of genetic resources, including by appropriate access to genetic resources and by appropriate transfer of relevant technologies, taking into account all rights over those resources and to technologies, and by appropriate funding, *thereby contributing to the conservation of biological diversity and the sustainable use of its components*.[105] (emphasis added)

The PIP Framework's objective is:

> ... to improve pandemic influenza preparedness and response, and strengthen the protection against the pandemic influenza by improving and strengthening the [GISRS], with the objective of a fair, transparent,

98 Fidler (n 43 above); Kelly Lee and David Fidler, 'Avian and pandemic influenza: progress and problems with global health governance' (2007) 2(3) Global Public Health 215–234.

99 Fidler (n 43 above) 90.

100 Ibid.

101 Noting that Fidler's argument is, of course, much more nuanced than what has been presented here.

102 Art 1. See also section on 'Contribution to sustainable development, as reflected in internationally agreed goals', page 437 below.

103 CBD Art 1. This objective continues: 'including by appropriate access to genetic resources and by appropriate transfer of relevant technologies, taking into account all rights over those resources and technologies, and by appropriate funding'.

104 Ibid 1.

105 Ibid.

> equitable, efficient, effective system for, on an equal footing: (i) the sharing of H5N1 and other influenza viruses with human pandemic potential; and (ii) access to vaccines and sharing of other benefits.[106]

Clearly, the overarching objectives of equity and fairness in the sharing of benefits associated with the use of genetic resources in the CBD, Nagoya Protocol and PIP Framework align. Notwithstanding the fact that isolating viruses from their hosts could be considered an act of *ex situ* biodiversity conservation, it cannot be said that the PIP Framework actively contributes to realising the biodiversity conservation and sustainable use objectives envisaged by the CBD and explicitly tied to ABS in the Nagoya Protocol. Nevertheless, we feel that not meeting this criterion is not essential to recognition as a specialized ABS instrument, so the relevant question here is to what extent the PIP Framework actually contributes to the fairness and equity objectives of the CBD and Nagoya Protocol, addressed below.

Fairness and equity in the sharing of benefits

The PIP Framework uses the term 'benefits' to refer to both the allocation of tangible diagnostics, vaccines and antivirals provided for specifically under the new PIP Framework and its SMTA2s,[107] and to those benefits that were already made available through the GISN (now GISRS). These include having the WHO coordinate pandemic preparedness and response,[108] and providing 'risk assessment and early warning information ... to all countries'.[109] Throughout the negotiation of the PIP Framework, it was clear that the countries demanding a fair and equitable share of the benefits arising from the utilisation of their viruses were referring to a fair share of the tangible diagnostics, vaccines and antivirals that had previously been denied them.[110] The value of the PIP Framework for these member states is that the pharmaceutical companies that were previously profiting from the use of their genetic resources were 'now requested to share the *actual benefits* of commercial research – be they in the form of vaccines, medical treatment, relevant production licences or similar means – on the basis of material transfer agreements'[111] (emphasis added).

The PIP Framework generates new benefits (as in, benefits that were not generated by the GISN prior to the existence of the PIP Framework) through two mechanisms: SMTA2s and an annual subscription payment from pharmaceutical companies using the GISRS, known as

106 Art 2.
107 Art 5(4)(2) and annex 2.
108 Art 6.
109 Arts 6(0)(2)(i) and 6(2).
110 WHO (n 37 above).
111 Wilke, 'The WHO's Pandemic Preparedness Framework' (n 4 above) 319.

Partnership Contributions.[112] The SMTA2 is signed by third parties that receive PIP biological materials from the GISRS, committing to provide benefits to the WHO like vaccines and antivirals in accordance with their abilities. These obligations are only enacted in the event of an influenza pandemic, when the WHO will distribute these benefits 'to developing countries, particularly affected countries, according to public health risk and needs'.[113]

The SMTA2 was a highly innovative mechanism under international law as it took the PIP Framework's provisions from its (soft law) WHA resolution into the realm of private law, where non-state actors (including pharmaceutical companies) could be regulated. This was seen as a step towards fair and equitable benefit sharing as it reduced the power dynamics between the (often) developing countries negotiating mutually agreed ABS terms with multinationals.[114] However, since concluded SMTA2s between vaccine manufacturers and the WHO have become publicly available, a more critical perspective has emerged doubting that the SMTA2s will deliver tangible benefits during an influenza pandemic.[115]

Since the adoption of the PIP Framework in 2011, 13 vaccine and antiviral manufacturers have entered into SMTA2s. The WHO has reported that these agreements would provide the agency with 400 million doses of pandemic influenza vaccine, 10 million treatment courses of antiviral drugs, 250,000 diagnostic kits, and 25 million syringes in the event of an influenza pandemic. The PIP Framework states that one-third of this stockpile 'will be for use in affected countries, according to public health risk and need, to assist in containing the first outbreak or outbreaks of an emerging pandemic';[116] and two-thirds 'will be for distribution, once a pandemic begins, to developing countries that have no or inadequate access to H5N1 influenza vaccines, on a per capita basis, with use to be determined by those countries'.[117]

It is important to note that the stockpile of 400 million doses of pandemic influenza vaccine is, by necessity, a virtual stockpile. The genetic drift of the virus is what necessitates the selection of

112 WHO, 'Pandemic Influenza Preparedness (PIP) Framework: Partnership Contribution (PC) Preparedness High-Level Implementation Plan II 2018–2023' Revised Version (WHO 2019) 3.

113 Art 6(0)(2)(iii), PIP Framework.

114 Wilke, 'The WHO's Pandemic Preparedness Framework' (n 4 above) 329; Nicole Jefferies, 'Levelling the playing field? Sharing of influenza viruses and access to vaccines and other benefits' (2012) 20(1) Journal of Law and Medicine 59.

115 Mark Eccleston-Turner, 'The Pandemic Influenza Preparedness Framework: a viable procurement option for developing states?' 17(4) Medical Law International (2017); Rourke (n 3 above).

116 Art 6(9((2) (i).

117 Art 6(9)(2) (ii).

new seasonal influenza vaccine strains each year, and genetic shifts make it impossible to determine what vaccine should be stockpiled in preparation for a pandemic. Thus, the PIP Framework's vaccine stockpile is on paper until such time as those pharmaceutical manufacturers that are party to the SMTA2s start producing pandemic vaccines and those SMTA2s are operationalised. Using the projections for the number of promised vaccine doses in the virtual stockpile, it is clear that it is unlikely to be sufficient to meet demand in developing countries during an influenza pandemic.[118] Indeed, at the point at which a vaccine becomes available during a pandemic (approximately six to seven months into the outbreak assuming no production problems),[119] demand in affected and developing countries that have no or inadequate access will far exceed the doses available through the PIP Framework's virtual stockpile.

Our key concern is that these benefits may not even materialise at all. In the first instance, the contractual obligations contained in the concluded SMTA2s between WHO and vaccine manufacturers may lack legal force.[120] One SMTA2 made publicly available by the WHO even included clauses that indemnified the manufacturer for almost every foreseeable difficulty to uphold its contractual obligations, including an influenza pandemic.[121] If the SMTA2s are upheld and manufacturers do donate a portion of their real-time vaccine production (this is just one example of the SMTA2 benefit-sharing options), there are still barriers to having those vaccines delivered to the WHO as promised:[122]

> ... concern has been expressed by the industry that during an influenza pandemic, member states with domestic [pandemic influenza vaccine] production within their territory would place restrictions on exports of [pandemic influenza vaccine] that have been committed to the PIP stockpile, until domestic demand had been fulfilled.[123]

Assuming, however, that the promised benefits will be available to the WHO, the PIP Framework provides no further guidance on how they are to be shared, other than to say:

> As regards the benefits outlined in this Framework, WHO should pay particular attention to policies and practices that promote the fair, equitable and transparent allocation of scarce medical resources

118 Eccleston-Turner (n 115 above).

119 WHO, 'Pandemic influenza vaccine manufacturing process and timeline: Pandemic (H1N1) 2009' Briefing Note 7, 6 August 2009.

120 Rourke (n 3 above) 100–103.

121 Ibid 102.

122 Ibid; Eccleston-Turner (n 115 above); Mark Eccleston-Turner, Alexandra Phelan and Rebecca Katz, 'Preparing for the next pandemic: the WHO's Global Influenza Strategy' (2019) 381(23) New England Journal of Medicine 2192.

123 Eccleston-Turner (n 115 above).

> (including, but not limited to, vaccines, antivirals and diagnostic materials) during pandemics based on public health risk and needs, including the epidemiology of the pandemic.[124]

There is no further guidance about how tangible benefits should be apportioned during a pandemic beyond the virtual stockpile distribution ratio of one-third for affected and two-thirds for developing countries detailed above.[125] It is important to note that diagnostics, vaccines and antiviral medicines promised under the SMTA2s will be provided to the WHO for onward transfer to affected or developing countries in real-time. This means that the WHO will have to make decisions about whether the first shipment is to go to affected country A or affected country B, and on what basis. The PIP Framework provides inadequate guidance on how this is to be determined, and it is therefore difficult to ascertain whether the PIP Framework does indeed meet the criteria of fair and equitable sharing for those benefits promised under the SMTA2 mechanism. It is an unfortunate fact that we will not be able to judge the equity and fairness criteria until such time as there is an influenza pandemic, and all of the PIP Framework's benefit-sharing provisions come into play.

The second mechanism through which the PIP Framework generates benefits is the annual Partnership Contribution. This is essentially an annual subscription payment to the WHO by any '[i]nfluenza vaccine, diagnostic and pharmaceutical manufacturers' that make use of the GISRS.[126] Unlike the benefit-sharing that occurs under the SMTA2s, benefit-sharing through Partnership Contributions is not contingent upon an active influenza pandemic. The annual Partnership Contributions are supposed to total the equivalent of half of the annual running costs of the GISRS and are used to improve pandemic preparedness and response.[127] The Partnership Contribution from vaccine, diagnostic and antiviral manufacturers generates USD28 million per year for the WHO, of which 10 per cent is used to fund the PIP Framework's governing body (the PIP Secretariat), and 'of the remainder, 30% are set aside for response during an influenza pandemic and 70% of funds are allocated for preparedness'.[128] As of the most recent Partnership Contribution report, the vast majority of that preparedness funding had been allocated to 'Laboratory and Surveillance Capacity Building',[129] outstripping the funding allocated

124 Art 6(1).
125 Art 6(9)(2).
126 Art 6(14)(3).
127 Ibid.
128 WHO, 'Pandemic Influenza Preparedness Framework Progress Report 1 January 2018–30 June 2019' (2019) vi.
129 Ibid 6.

to the categories of Regulatory Capacity Building, Risk Communications and Community Engagement, Burden of Disease, Influenza Pandemic Preparedness Planning and Planning for Deployment combined.[130] The funding for Laboratory and Surveillance Capacity Building, as the name suggests, is aimed at strengthening the core activities of the GISRS. Capacity-building activities under this banner include '[s]trengthen[ing] data and information sharing from national to regional and global platforms and improv[ing] data management systems' and '[f]acilitat[ing] influenza sample shipment to GISRS by providing necessary consumables and train[ing] NIC staff to select and ship quality samples'.[131]

We are not suggesting for a moment that these are not vital activities, and we fully support the strengthening of the GISRS. However, from an ABS point of view (and this is, of course, the purpose of this exercise), the benefits provided through the Partnership Contributions (increasing lab and surveillance capacity, NIC training and community engagement) are not the type of benefits that developing countries had in mind as the PIP Framework was being negotiated. The access part of the transaction was supposed to be about member states providing their sovereign viruses to the WHO (GISRS), and the benefit-sharing portion was supposed to be about the fair and equitable distribution of diagnostics, vaccines and antivirals. The GISRS capacity-building activities funded by the Partnership Contributions can ultimately be seen as strengthening the access side of the ABS transaction.[132]

Because the tangible benefits of the PIP Framework cannot 'be provided in a preferential manner to the country from which the virus had originated, but rather as a pooled benefits system based on public health risk and need (aimed at developing countries)',[133] there is no direct link between the access and the benefit-sharing side of the PIP Framework's ABS transaction. This would not necessarily be a bad thing if all resources (virus samples and benefits) were treated as common pools.[134] But under the PIP Framework, the viruses and associated data contributed to GISRS by WHO member states continue to be treated as global public goods[135] (and strengthening the GISRS increases the value of those public goods), but the fair and equitable sharing of diagnostics, vaccines and antivirals is in no way guaranteed

130 Ibid.
131 Ibid.
132 See also Rourke (n 3 above).
133 Ibid.
134 See Wilke, 'The WHO's Pandemic Preparedness Framework' (n 4 above).
135 Youde (n 32 above) 115.

by the PIP Framework, meaning that these continue to be treated as private goods and are likely to be distributed accordingly.[136]

Legal certainty with respect to access to genetic resources or traditional knowledge and to benefit-sharing

The PIP Framework applies only to 'PIP biological materials', which are defined by the PIP Framework as 'human clinical specimens, virus isolates of wild type … influenza viruses with human pandemic potential; and modified viruses … with human pandemic potential'.[137] However, no further definition is given as to what influenza virus subtypes constitute having 'human pandemic potential' for the meaning of the PIP Framework. It therefore falls to individual NIC laboratories to make a determination if the virus samples they hold ought to be shared with the GISRS under the terms of the PIP Framework. Often NICs do not have the ability to conduct the testing required to determine whether a virus sample has human pandemic potential, testing that occurs at the regional CCs. Therefore, samples are often transferred from the NICs to the CCs before it is clear whether the PIP Framework even applies to those samples.[138]

There is also a level of legal uncertainty with respect to third-party transfers of PIP biological materials. The host country NICs transfer virus samples to regional CCs under the SMTA1, and the act of providing samples constitutes consent to the onward transfer of their PIP biological materials to other GISRS laboratories.[139] It also constitutes consent for the CCs to transfer PIP biological materials to third-parties under an SMTA2.[140] An SMTA2 authorises the transfer of the PIP biological materials from GISRS to parties that sit outside of the GISRS network, including academic laboratories and research institutes, as well as diagnostic and vaccine manufacturers.[141] As already outlined, it is through the SMTA2 that these third-parties agree to provide benefits in return for access to PIP biological materials. Despite being a standardised agreement, the SMTA2s can take a long time to negotiate. Accordingly, the WHO has deemed it sufficient to provide PIP biological materials to third parties using only a shipping notice that contains a weak and uncertain '[a]greement to conclude' an SMTA in some undefined future.[142] This shipping notice states that 'by

136 See also Eccleston-Turner (n 115 above).
137 Art 4(1).
138 See Rourke (n 3 above) 100.
139 Art 5(4)(1) and annex 1.
140 Art 5(4)(2) and annex 2.
141 Ibid.
142 WHO, 'PIPBM Shipping Notice' (1 October 2019).

receiving these materials you are signalling your intention to be bound by the terms of a future SMTA2'. The legal effect of this is uncertain, but clearly 'vague references to unknown terms cannot form the basis of a contract'.[143] Significant elements of the SMTA2 are negotiable, not just for the benefit-sharing commitments manufacturers agree to, but also provisions on liability and indemnity, jurisdiction and shipping arrangements. As such, it is reasonable to assume that third-party recipients may not be able to conclude an SMTA2 with the WHO on mutually agreed terms, or in good time. It is unclear what this would mean in terms of tangible benefit-sharing in the event of a pandemic, or what may eventuate if the third party and the WHO cannot reach an agreement that forms the basis of an SMTA2 and that third party had already used the PIP biological materials before concluding an SMTA2. The lack of specificity and binding nature of the terms of the WHO's shipping notice for PIP biological materials cannot be said to provide any legal certainty over transfers of PIP biological materials to third parties without an SMTA2. Given the extant legal uncertainty in both the access and the benefit-sharing sides of the ABS transaction under the PIP Framework, it fails to meet this criterion of the specialized instrument Study.

Contribution to sustainable development, as reflected in internationally agreed goals

The Study highlights 'the explicit link established between benefit-sharing and the other two objectives of the CBD – *conservation and sustainable use*' (emphasis in original) and that ABS 'is not to be pursued in isolation from the broader framework established by the CBD'.[144] It is hard to interpret the ABS provisions of the PIP Framework, or indeed any part of the PIP Framework, as contributing to 'inter alia, the selection and management of protected areas and species, the restoration of degraded ecosystems and the protection and promotion of traditional knowledge',[145] again, notwithstanding the fact that the isolation of wild-type influenza viruses and their storage in laboratory freezers might be considered species conservation. The Preamble of the Nagoya Protocol presents a broad conception of sustainable development, which includes the 'contribution to sustainable development made by technology transfer and cooperation to build research and innovation capacities for adding value to genetic resources in developing countries' and the 'potential role' of ABS to

143 Aaron Perzanowski and Jason Schultz, *The End of Ownership: Personal Property in the Digital Economy* (MIT Press 2016) 68.

144 CBD (n 4) 8.

145 Ibid.

contribute 'to achieving the Millennium Development Goals'.[146] The Millennium Development Goals were succeeded by the Sustainable Development Goals (SDGs), adopted by all UN member states in 2015.[147] For the PIP Framework, the SDG 3 to ensure healthy lives and promote well-being for all at all ages is particularly relevant.[148]

The SDG 3 targets include 'access to safe, effective, quality and affordable essential medicines and vaccines for all', 'support[ing] the research and development of vaccines and medicines' and 'strengthen[ing] the capacity of all countries, in particular developing countries, for early warning, risk reduction and management of national and global health risks'.[149] While we have concerns that the PIP Framework will be unable to deliver the promised vaccines during an influenza pandemic, it is clear that the PIP Framework's Partnership Contributions that go towards strengthening GISRS capacity undoubtedly contribute to sustainable development under SDG 3. Thus, the PIP Framework could be said to meet this criterion of the Study.

Other general principles of law including good faith, effectiveness and legitimate expectations

It is possible to separate out the access side of the PIP Framework and the benefit-sharing side when determining whether it meets the general principle of effectiveness. While the non-binding PIP Framework may have worked to codify virus-sharing norms, there is nothing in the PIP Framework that can compel countries to share their influenza viruses with the GISRS, or avert a similar crisis to the one that was the very impetus for the PIP Framework.[150] Thus, the PIP Framework is only effective as an access mechanism in as much as countries providing virus samples trust that the benefit-sharing mechanisms will be effective.

The effectiveness of the PIP Framework's benefit-sharing mechanisms is difficult to ascertain, as the full complement of benefit-sharing options has not been put to the test during an influenza pandemic. Unfortunately, the PIP Framework 'has not sufficiently engaged with, or found appropriate solutions to, the current market-

146 Preamble.

147 United Nations Development Programme (UNDG), SDGs.

148 Michala Hegermann-Lindencrone et al, 'Innovative Pandemic Influenza Preparedness framework paves the way for sustainable improvements to pandemic preparedness' (2018) 4(1) Public Health Panorama 79.

149 UNDG, SDG 3: Ensure healthy lives and promote well-being for all at all ages.

150 'Little that triggered or transpired during [the 2006/2007 virus sharing] controversy would therefore violate the framework that supposedly resolves it': Smith (n 43 above).

based structural hurdles that prevent equitable access to vaccines, namely limited overall global production capacity, the prevalence of [advanced purchase agreements] and the need for more private sector investment'[151] and is highly unlikely to be an efficient or effective benefit-sharing tool in the event of a pandemic.

Another challenge affecting the effectiveness of the PIP Framework, both in terms of enabling access and for the sharing of benefits, is synthetic biology and the move toward using influenza genetic sequence data for developing vaccines. In 2016, the PIP Framework Review Group noted that genetic sequence data (GSD) could in some cases be used instead of physical virus samples during pandemic risk assessment and for vaccine development,[152] and the technical developments are such that this move away from using physical viral samples and instead using GSD is expected to grow.[153] This means that third-party users of 'dematerialised' PIP biological materials can avoid entering into an SMTA2 with the WHO despite still benefiting from the contribution of member states and the outputs of the GISRS.[154] The WHO has stated that, while there is no obligation to enter into an SMTA2, any manufacturers that use 'GSD produced by the GISRS … are expected to contribute an annual Partnership Contribution payment'.[155] What this fails to appreciate, however, is that it is usually possible to access this GSD free of charge on any number of publicly accessible databases, often without any way of determining who has accessed that data (that is, you do not need to pay a Partnership Contribution to see or use GSD from influenza viruses with human pandemic potential).

While the PIP Framework recognises 'that in some instances the publication of [GSD] has been considered sensitive by the country providing the virus',[156] it still encourages all member states to share GSD,[157] despite the fact that it is not included in the PIP Framework's ABS regime. This loophole has been acknowledged by the WHO since 2013, and in 2014 the PIP Framework's Technical

151 Adam Kamradt-Scott and Kelly Lee, 'The 2011 Pandemic Influenza Preparedness Framework: global health secured or a missed opportunity?' (2011) 59(4) Political Studies 831, 844–845.

152 WHO, 'Review of the Pandemic Influenza Preparedness Framework for the Sharing of Influenza Viruses and Access to Vaccines and Other Benefits: Report of the 2016 Pandemic Influenza Preparedness Framework Review Group' (18 November 2016).

153 See WHO, 'Fact sheet: new technologies using genetic sequence data', 4 April 2018.

154 WHO, Approaches to Seasonal Influenza and Genetic Sequence Data under the PIP Framework, 14 December 2018, 17.

155 WHO (n 56 above) 18.

156 Art 5(2)(3).

157 Art 5(2)(1).

Expert Working Group convened to assess the scientific, technical, operational and intellectual property implications of using GSD instead of physical viruses for research and vaccination production, and how the use of this data could be monitored. The 2016 meeting of the PIP Framework's Technical Working Group on GSD proposed amending the PIP Framework to include GSD within the definition of PIP biological materials and therefore include GSD utilisation in the ABS arrangements of the PIP Framework.[158] However, none of these proposals has come to pass, and to date no amendments or conclusions have been reached to minimise the impact that free and open access to influenza GSD will have on the sharing of physical influenza samples under the PIP Framework.

On the point of general principles of law, the extent to which dispute resolution under the PIP Framework aligns with principles of good faith, justice and fairness is also questionable. The PIP Framework does provide for dispute resolution for member states providing PIP biological materials under the SMTA1.[159] In the first instance, parties are to attempt to settle the dispute via negotiation. However, in the event of this failing, 'one of the parties concerned may refer the dispute to the Director-General, who may seek advice of the Advisory Group with a view to settling it'.[160] This seems logical if the dispute to be resolved is between the party providing PIP biological materials and the party using those materials to generate benefits to be shared with the provider. However, what this dispute resolution mechanism fails to acknowledge is that there is no agreement between the provider and user parties under the PIP Framework. The SMTA1 is an agreement between the member states providing PIP biological materials and the WHO, and the SMTA2 is an agreement between WHO and a third-party user. This means that the WHO is a party to both the SMTA1 and the SMTA2, and there is no direct link between member states and third parties like vaccine manufacturers. It seems rather bizarre to have designed a dispute resolution mechanism which is to be adjudicated upon by one of the parties to the agreement under dispute, nor does it seem like a system that would meet legitimate expectations of fair and effective dispute resolution.

158 This was also outlined as a potential option in WHO (n 56 above) 30; GISAID (n 56 above) 24–26.

159 Annex 1, art 7.

160 Annex 1, art 7(2).

THE PIP FRAMEWORK AND LEGAL CERTAINTY IN ABS ARRANGEMENTS

The previous section examined aspects of the PIP Framework against the nine criteria for specialized ABS instruments outlined in the 2018 Study for the Subsidiary Body on Implementation to the CBD. Despite some shortcomings, including the fact that the PIP Framework does not consider virus-related traditional knowledge of Indigenous Peoples and local communities, we are satisfied that the PIP Framework likely meets the four criteria for specialization. However, the PIP Framework has major shortcomings when it comes to realising three of the five criteria on supportiveness: fairness and equity in benefit-sharing, creating legal certainty for ABS and the general legal principles of effectiveness and legitimate expectations, which all link back to legal certainty.

Despite applying only to influenza viruses with human pandemic potential, the PIP Framework reinforces the norm of sharing both seasonal and pandemic influenza viruses. This is because countries often do not know whether the samples they are providing to the GISRS are classified as seasonal or pandemic influenza viruses until the analysis occurs within the GISRS. This means that, for whatever legal certainty the PIP Framework might provide for the transfer of pandemic influenza viruses to third parties outside of GISRS laboratories, there is no equivalent certainty for the seasonal influenza viruses that have already been contributed to the GISRS by member states. These are still subject to benefit-sharing obligations under the CBD and the CBD/Nagoya Protocol ABS regimes, but the member states have lost any ability to monitor or functionally control the use of seasonal influenza virus samples that they have already provided to the GISRS. Indeed, vaccine manufacturers often obtain their seasonal candidate vaccine virus strains from the GISRS, and there has already been confusion about whether the PIP Framework or the Nagoya Protocol should govern such transfers.[161] Countries that routinely contribute viruses to the GISRS guarantee the WHO access to their seasonal *and* pandemic influenza viruses but could be doing so on the promise of benefits that are linked solely to the pandemic influenza viruses. Put simply: countries are generally expected to provide access to all influenza viruses but are promised benefits in return for just a minority of them. Whether or not those benefits will be forthcoming is

161 WHO (n 56 above) 30. Given the PIP Framework is applicable only to those pandemic influenza viruses shared with the GISRS, the Nagoya Protocol is the instrument that governs transfers of seasonal influenza candidate vaccine strains to third parties. This means that those countries can negotiate prior informed consent and mutually agreed terms on a bilateral, case-by-case basis.

another level of legal uncertainty that sovereign nations must assume when providing influenza viruses under the PIP Framework.

The issue of GSD presents yet another access-related problem that undermines the ability of the WHO to secure benefits through the PIP Framework and therefore any legitimate expectations of receiving such benefits by those providing samples to the GISRS. As outlined above, GSD is not included in the definition of PIP biological materials, and synthetic biology technology has developed such that many uses of influenza viruses by pharmaceutical companies and other third-party users no longer necessitate access to physical samples.[162] The WHO has claimed that users of GSD generated through the GISRS would still be expected to make an annual Partnership Contribution payment,[163] but this completely disregards the fact that similar (or indeed identical) GSD is usually available through open access databases like GenBank, and that there is no reliable method for tracking or tracing the use of such data. While GSD is often framed as an emerging issue, this has been on the radar of the WHO since the PIP Framework's inception,[164] and, as yet, there has been no decision on how to deal with this loophole despite it severely undermining the legal clarity that the PIP Framework is supposed to provide for those countries contributing physical samples to the GISRS.

As an ABS instrument, the PIP Framework does not adequately define which actions constitute access and which it considers benefit-sharing. For instance, the provision of candidate vaccine viruses to influenza vaccine manufacturers should be considered providing access to PIP biological materials. However, the PIP Framework addresses this under article 6 on benefit-sharing. That is, the PIP Framework treats the provision of some PIP biological materials to vaccine manufacturers as benefit-sharing. These are clearly not the benefits for which the PIP Framework was intended to ensure fair and just distribution. Any future specialized instruments would do well to indicate precisely what actions constitute access and which constitute benefit-sharing for the purposes of providing legal clarity.

The PIP Framework does specify a limited number of benefits which are directly tied to providing PIP biological materials: the

162 Michelle Rourke et al, 'Policy opportunities to enhance sharing for pandemic research' (2020) 368(6492) Science 716; Michelle Rourke, Alexandra Phelan and Charles Lawson, 'Access and benefit-sharing following the synthesis of horsepox virus' (2020) 35(8) Nature Biotechnology 539

163 WHO (n 56 above) 18.

164 Intergovernmental Meeting on Pandemic Influenza Preparedness: Sharing of Influenza Viruses and Access to Vaccine and Other Benefits, Sharing of Influenza Viruses and Access to Vaccines and Other Benefits: Interdisciplinary Working Group on Pandemic Influenza Preparedness, Report by the Director-General, A/PIP/IGM/4.

active participation of scientists from originating laboratories in scientific projects associated with those materials;[165] access to the genetic sequence data and analyses derived from those materials;[166] and acknowledgment of 'the contribution of collaborators' in downstream 'presentations and publications'.[167] But many countries are contributing to the GISRS on the promise of tangible benefits to be delivered in the event of an influenza pandemic: those diagnostics, antiviral medications and vaccines that can help the worst-hit countries cope with the crisis. The SMTA2s – the mechanism used by the PIP Framework to generate these tangible benefits for distribution by the WHO in the event of a pandemic – are as yet untested and appear unlikely to deliver the quantum of benefits required to adequately respond to an influenza pandemic, or even those envisaged in the PIP Framework itself. Furthermore, there is little legal certainty as to how these tangible benefits will be distributed as the PIP Framework simply states that the WHO will distribute benefits 'according to public health risk and needs'.[168] This is assuming that these tangible benefits are indeed available for distribution, which depends in large part on whether the PIP Framework's SMTA2s are a viable legal instrument for securing such benefits.[169]

The point of benefit-sharing for an instrument designed to create a common pool of resources, like that created by the PIP Framework,[170] is that it incentivises countries to provide access to their sovereign genetic resources. Access and benefit-sharing is a transactional mechanism; a *quid pro quo*. For now, physical virus samples are required to manufacture vaccines. The bilateral version of this transaction is envisaged as the vaccine manufacturers willing to pay (share benefits with) the country that can provide access to the raw materials required to make their product. Countries can choose to regulate this transaction through their own legislative, administrative or policy measures, or can choose to have the transaction facilitated through the WHO under the PIP Framework. If countries determine that the sharing of benefits from the PIP Framework's common pool is not fair and equitable or will not be forthcoming, then they are unlikely to continue to provide access to the viruses. Thus, there is an inherent tension built into the way that the PIP Framework proposes to allocate benefits. If the WHO provides PIP Framework benefits based solely on public health needs and irrespective of whether a particular country has

165 Annex 1, art 5(2).
166 Art 5(2)(1).
167 Annex 1, art 5(3).
168 Art 6(0)(2)(iii).
169 See Rourke (n 3 above); Eccleston-Turner (n 115 above).
170 See Wilke, 'The WHO's Pandemic Preparedness Framework' (n 4 above) 325.

contributed virus samples to the GISRS, then there is no incentive for any individual country to provide its viruses. The GISRS will continue to operate without such countries' contributions, and they still stand to receive benefits if warranted on the basis of public health needs.[171] But, if the WHO decides that it must prioritise the delivery of benefits to countries that have continually provided viruses to the GISRS in order to keep the PIP Framework's incentive structure strong, it has now entered dangerous territory, acting on political expediency rather than on the basis of public health alone, and therefore outside of the PIP Framework's expressed provisions on benefits distribution. This highlights the folly of using the ABS transaction as a means of securing access to resources that are absolutely essential to global health security. If the PIP Framework does not get benefit-sharing right, then the WHO risks continued access to the influenza viruses (seasonal and pandemic) that the world needs to monitor and respond to seasonal influenza, detect and alert the world to a potential pandemic, and ensure the samples required by vaccine manufacturers to help respond to that pandemic are available immediately.

The Study's interest in legal certainty (criterion 7) referred to 'legal certainty with respect to access to genetic resources ... and to benefit-sharing'. We could refer to this as internal certainty (the interaction between the parties within the PIP Framework). However, there is further legal uncertainty that must be addressed about the PIP Framework's relationship with the Nagoya Protocol (and the CBD) if it were to be recognised as a specialized instrument, or, external certainty (interactions with other international instruments and norms). In addressing the PIP Framework's relationship to the Nagoya Protocol, the 2016 PIP Framework Review Group stated that the recognition of the PIP Framework as a specialized instrument:

> ... should facilitate fulfilment of the PIP Framework's access and benefit sharing objectives by ensuring that all countries would handle IVPP [influenza viruses with human pandemic potential] in the same way. IVPP access and sharing would be covered for Nagoya Protocol purposes by the PIP Framework, and therefore not require bilateral agreements on a case-by-case basis.[172]

This sentiment was repeated in a 2017 WHO study into the public health implications of the Nagoya Protocol which stated that recognition of the PIP Framework as a specialized instrument:

> ... would mean that the Nagoya Protocol's requirements for case-by-case Prior Informed Consent and Mutually Agreed Terms would not

171 This is analogous to the 'Tragedy of the Commons': Garrett Hardin, 'The Tragedy of the Commons' (1968) 162(3859) Science 1243.

172 WHO (n 18 above) 96.

> apply with respect to influenza viruses with human pandemic potential. This could promote 'legal certainty' with respect to such pathogens, strengthening the mechanisms of the PIP Framework.'[173]

The PIP Framework Review Group and PIP Secretariat[174] appear to be working on the assumption that recognition as a specialized instrument would mean that 'all countries would handle IVPP in the same way', that is, the PIP Framework way.[175] The assumption of external legal certainty upon recognition as a specialized instrument is inaccurate. The PIP Framework 'only functions as a specialized ABS instrument for influenza viruses where transfers are covered by the Framework's binding contract clauses'.[176] Wilke states that 'the [PIP] Framework may only partially be considered a specialized ABS instrument within the meaning of the [Nagoya] Protocol',[177] because it does not include *all* pandemic influenza viruses with human pandemic potential, just the ones that countries have chosen to share under the terms of the PIP Framework's SMTAs. There are still influenza viruses with human pandemic potential shared bilaterally outside of the GISRS. That means that 'the Nagoya Protocol must remain applicable in the background',[178] and, thus, not all influenza viruses with human pandemic potential can be 'considered exempt from the Nagoya Protocol's scope by virtue of Article 4.4'.[179] Thus, recognition of the PIP Framework as a specialized instrument achieves nothing in the way of clarifying the legal confusion surrounding the application of the PIP Framework; confusion that has already caused delays in accessing influenza viruses for vaccine production.[180]

CONCLUSION

Article 4 of the Nagoya Protocol affords parties the latitude needed to design and implement specialized ABS arrangements for particular subsets of genetic resources that are ill-suited to the default bilateral measures envisaged in the CBD and Nagoya Protocol. There have been calls to have the PIP Framework recognised as a specialized instrument

173 WHO, Implementation of the Nagoya Protocol and Pathogen Sharing: Public Health Implications (WHO 2017) 23.

174 National Academies of Sciences, Engineering, and Medicine, *Exploring Lessons Learned from a Century of Outbreaks: Readiness for 2030: Proceedings of a Workshop* (2019) 96.

175 WHO (n 18) 22–23.

176 Wilke, 'A healthy look at the Nagoya Protocol' (n 4 above) 126.

177 Ibid 126.

178 Ibid 145.

179 Ibid 146.

180 WHO (n 56 above) 30.

with the assertion that doing so would clarify the applicability of the PIP Framework in relation to the CBD and Nagoya Protocol and therefore the appropriate global ABS arrangements for influenza viruses.[181] The analysis above indicates that recognition as a specialized instrument would not result in the 'disapplication'[182] of the CBD/Nagoya Protocol for all influenza viruses with human pandemic potential: only for those pandemic influenza virus samples that countries choose to share with the WHO under the terms of the PIP Framework's SMTAs. The confusion will remain even if the PIP Framework is formally recognised as a specialized international ABS instrument.

This article has not touched on the potential process for recognition of a specialized instrument, but rather focused on the form that a specialized instrument would need to take in order to qualify for that recognition. But the mechanics of recognition are important. The Meeting of the Parties to the Nagoya Protocol is likely the only body that could possibly recognise a specialized instrument because the specialized instrument provision is only found in the Nagoya Protocol. There is no equivalent provision in the CBD, so states parties to just the CBD are not bound to recognise the existence of any specialized instrument. It is not clear whether recognition would mean that the 123 contracting parties to the Nagoya Protocol would have to treat the PIP Framework as a specialized instrument under their domestic ABS legislation while the remaining 73 states parties to the CBD alone (as well as the United States and the Holy See which are party to neither agreement) would not have to accept the PIP Framework's new found status as a specialized instrument at all. Thus, recognition of the PIP Framework as a specialized ABS instrument would not necessarily change anything for more than a third of WHO member states.

All of this indicates that recognition as a specialized instrument would be a purely symbolic (perhaps political) gesture, not able to alter the legal status of the PIP Framework. That is not to say that symbolic gestures are unimportant. We have made this point before, and we want to reiterate it here: access to viruses and other pathogen samples need not be connected to the sharing of vital medicines and vaccines. These issues are both public health issues, but they do not need to be linked through the ABS transaction. The PIP Framework crystallised these separate issues as a single ABS issue, and ABS has its home in the UN system with the CBD and Nagoya Protocol. This is potentially why recognition as a specialized instrument is so appealing for the PIP Secretariat and the WHO more broadly. It seems like an easy fix. If the PIP Framework is considered a specialized ABS instrument under the

181 See eg WHO (n 173 above) 9.

182 CBD (n 4 above) 2.

Nagoya Protocol, then the WHO remains relevant in the discussions about pathogen sample sharing and the Nagoya Protocol ceases to present a conceptual challenge to the PIP Framework, one where the ultimate goals of both instruments are potentially at odds (conservation of genetic resources versus eradication of disease). It also means the PIP Framework would have the international ABS stamp of approval and an endorsement of the suitability of the transactional mechanism for the sharing of pathogens *and* the sharing of pharmaceuticals, and the WHO would no longer have to grapple with the unpalatable reality that it is encouraging countries to use their pathogen samples as currency to purchase the life-saving vaccines and medicines to which they should already have access.

Northern Ireland
Legal Quarterly
Autumn Vol. 72 No. 3 (2021) 448–479
Article DOI: 10.53386/nilq.v72i3.715

‘Survival of the fittest’? Perceptions of wellbeing at the Bar of Northern Ireland

Neil Graffin

The Open University

Emma Jones

University of Sheffield
Correspondence email: neil.graffin@open.ac.uk

ABSTRACT

This article, based on a qualitative study conducted in June–July 2019, assesses how barristers at the Bar of Northern Ireland perceive wellbeing and mental ill-health within their profession. It will argue that the Bar can be a competitive and potentially hostile workplace environment, leading to detrimental impacts on wellbeing. It will also contend that being a barrister in Northern Ireland provides its own unique challenges for practitioners due to the self-employed independent nature of the role, where practitioners do not work in chambers or stables like their counterparts in England & Wales and Scotland. At the same time, barristers spoke positively about the flexibility of their roles and beneficial forms of collegiality, as well as an evolving culture which places greater emphasis on wellbeing. This article will argue, using the ‘job demands and resources’ model, that efforts should be made to decrease job demands and increase the job resources of barristers of the Bar of Northern Ireland to improve levels of wellbeing. This model could also be applied to the Bar in other jurisdictions to assess the impact of both shared and unique challenges and opportunities.

Keywords: wellbeing; legal profession; the Bar of Northern Ireland; barrister; mental health; job demands; job resources.

INTRODUCTION

> You couldn’t sustain a practice and have mental health problems. The profession … you wouldn’t survive in it. It couldn’t work. (Barrister 6)

In recent years, the issue of wellbeing within the legal profession has become of increasing concern both internationally and within the United Kingdom (UK), with a growing body of evidence that lawyers experience higher levels of mental health issues and lower levels of

wellbeing than the general population.[1] A range of factors relating to the legal workplace have been implicated as potential causes, often including structural and cultural issues. For example, the high billing targets set by law firms, resulting in cultures of long working hours,[2] or the need to appear continually productive, which may be attributed to neoliberal dogma which promotes productivity at all costs, and often ignores the psychological wellbeing of practitioners.[3] Other issues identified include poor management or a lack of control over workload,[4] or an inability to achieve an appropriate work–life balance.[5]

The growing academic discipline regarding lawyer wellbeing has seen a proliferation of publications in the field, as well as the establishment of an international research group: Advancing Wellness in Law. Most academic research has been undertaken within the jurisdictions of the United States and Australia, with the wellbeing of legal professionals in the UK overlooked until recently. Work undertaken by Collier[6] sought to explore issues of wellbeing, gender and legal practice,[7] and a recent book by Jones et al,[8] based on a qualitative study of 30 practitioners, sought to establish perceptions of lawyer mental health and wellbeing from across the legal profession. Outside of academia in the UK, however, there has been striking media coverage of issues pertaining

1 Patrick R Krill, Ryan Johnson and Linda Albert, 'The prevalence of substance use and other mental health concerns among American attorneys' (2016) 10 Journal of Addiction Medicine 1, 46; Norm Kelk, Georgina Luscombe, Sharon Medlow and Ian Hickie, *Courting the Blues* (Brain & Mind Research Institute, University of Sydney, with Tristan Jepson Memorial Foundation 2009); Positive, Wellbeing at the Bar. A Resilience Framework Assessment (Bar Council 2015).

2 Adele J Bergin and Nerina L Jimmieson, 'Australian lawyer well-being: workplace demands, resources and the impact of time-billing targets (2014)' 21 Psychiatry, Psychology and Law 3, 427.

3 Margaret Thornton, 'Squeezing the life out of lawyers: legal practice in the market embrace' (2016) 25 Griffith Law Review 471.

4 Colin James, 'Lawyers' wellbeing and professional legal education' (2008) 42 The Law Teacher 85

5 Margaret Thornton, 'Work/life or work/work? Corporate legal practice in the twenty-first century' (2016) 23 International Journal of the Legal Profession 13.

6 Richard Collier, '"Love law, love life": neoliberalism, wellbeing and gender in the legal profession – the case of law school' (2014) 17 Legal Ethics 202; Richard Collier, 'Wellbeing in the legal profession: reflections on recent developments (or, what do we talk about, when we talk about wellbeing?)' (2016) 23 International Journal of the Legal Profession 41.

7 Collier, 'Love law, love life' (n 6 above).

8 Emma Jones, Neil Graffin, Rajvinder Samra and Mathijs Lucassen, *Mental Health and Wellbeing in the Legal Profession* (Bristol University Press 2020).

to lawyer wellbeing.[9] In addition, regulatory bodies have recognised the importance of wellbeing,[10] and studies have been conducted by the Junior Lawyers Division of the Law Society in assessing wellbeing of members.[11] The International Bar Association,[12] and LawCare[13] have also recently embarked on large-scale projects to assess lawyer wellbeing.

To date, much of the research focus has been on lawyers employed within areas of private practice, commonly within large law firms.[14] However, the legal profession itself is not a homogeneous grouping, and it is becoming increasingly apparent that wellbeing issues may manifest themselves in different ways and have differing causes and consequences within different populations.[15] The role of a barrister is one that has previously been identified as having specific challenges in terms of wellbeing.[16] It is a role unique to common law jurisdictions. Barristers form a body of regulated specialist legal advisers who are commonly self-employed and who provide a range of services, including 'advocacy and representation in court' as well as offering 'written advice, negotiation and mediation'.[17] At the Bar in Northern Ireland, there are around 600 barristers in total,[18] with only 20 graduates per year admitted to the Institute of Professional Legal Studies Bar training course (referred to colloquially as the Institute and based at Queen's

9 Elizabeth Rimmer, 'The harsh reality of a long hours culture – why Weil's April fool was no laughing matter' *Legal Week* (24 April 2015); Neil Graffin et al, 'The legal profession has a mental health problem – which is an issue for everyone' (*The Conversation*, 18 April 2019); The University of Law, 'How the legal world is changing its attitudes to mental health and wellbeing' *The Guardian* (London 14 October 2019).

10 See Solicitors Regulation Authority, 'Your health, your career' (SRA, 6 April 2020); Law Society of Scotland, 'LawScot Wellbeing'; Law Society of Ireland, 'Shrink me'.

11 Law Society Junior Lawyers Division, *Resilience and Wellbeing Survey Report 2019* (Law Society. 2019).

12 International Bar Association, 'Mental wellbeing in the legal profession'.

13 LawCare, 'Life in the law'.

14 Vivien Holmes, Tony Foley, Stephen Tang and Margie Rowe, 'Practising professionalism: observations from an empirical study of New Australian lawyers' (2012) 15 Legal Ethics 1, 29.

15 Jones et al (n 8 above).

16 Positive (n 1 above); Lloyd C Harris, 'The emotional labour of barristers: an exploration of emotional labour by status professionals' (2002) 39 Journal of Management Studies 4, 553.

17 Bar of Northern Ireland, 'What do barristers do?' (Bar of Northern Ireland 2020).

18 Bar of Northern Ireland, 'The Bar of Northern Ireland'.

University Belfast),[19] which is the local vocational prerequisite for joining the Bar of Northern Ireland (barristers can train elsewhere before being admitted).

The Bar in Northern Ireland is an almost century-old institution created after the partition of Ireland, when the first Inn of Court for Northern Ireland was established.[20] The Honorable Society of the Inn of Court of Northern Ireland is the body which governs the education, training and admittance of barristers in Northern Ireland. There are different routes to becoming a barrister in Northern Ireland, dependent on where a candidate studies, qualifies and completes training. For those studying in Northern Ireland, prospective candidates are required to complete a qualifying law degree, a Postgraduate Diploma in Professional Legal Studies at the Institute and receive a call to the Bar of Northern Ireland, where they then complete a 12-month pupillage.[21]

The Bar Council is responsible for the governance, regulation and representation of the profession. It is an elected body of 20 practising barristers, whose powers and functions are defined in the Constitution of the Bar of Northern Ireland and byelaws of the General Council of the Bar of Northern Ireland.[22] The Bar Council regulates all practising barristers through the Professional Conduct Committee. Each barrister is subject to the Code of Conduct of the Bar of Northern Ireland, which sets out standards of professional conduct and practice required of barristers.

The benchers of the Inn of Court are responsible for the admission of barristers to practise at the Bar. This includes the application process for call to the Bar, whether that is on a permanent or temporary basis, or by a transferred call from another jurisdiction. Benchers are senior lawyers drawn from the Bench and the Bar and include the Lord Chief Justice and Attorney General for Northern Ireland.[23]

In general, every barrister in independent practice will be a member of the Bar Library in Northern Ireland, although temporary admission may be granted to barristers from England & Wales who meet the required criteria.[24] The Bar Library provides a physical working space, with desks and computers, conferencing rooms, and other facilities (although barristers may choose instead to work from home).

19 Queen's University Belfast Institute of Professional Legal Studies, Information Booklet for Applicants for commencement in September 2020 of: The Bar Course at the Institute of Professional Legal Studies and the Solicitor Course at the Institute of Professional Legal Studies (QUB 2020) 16.

20 Ibid.

21 Ibid

22 Bar Library of Northern Ireland, 'Governing bodies'.

23 Ibid.

24 Section 20 of the Admission Rules of the Honorable Society of the Inn of Court of Northern Ireland.

The position and role of barristers in Northern Ireland is distinct from that of barristers elsewhere. Unlike barristers practising elsewhere in the UK, barristers at the Bar of Northern Ireland do not belong to chambers (common in England & Wales) or stables (common in Scotland). Instead, the majority conduct their work as self-employed sole traders with no administrative assistance from barrister's clerks. Clerks reside in chambers or stables in the other jurisdictions and undertake several tasks for barristers, such as keeping their diaries up to date, allocating and overseeing work arriving in, liaising between solicitors, clients and their barristers and collecting pay, *inter alia*.[25] These tasks are all required to be undertaken by the vast majority of barristers in Northern Ireland.

Clerks can also provide a counselling role, in the sense that they can support barristers in guiding them to areas of work appropriate to their talents. As Flood suggests, 'given the anxieties of barristers and their need for reassurance … [they can] adjust their expectations in subtle ways … [If] a barrister is not a robust advocate in court, the clerk can suggest a move towards advisory work'. This suggests that the role of the clerk can be one of career guide and advisor.[26] In addition, Flood discusses how clerks traditionally have provided a listening ear to barristers for their personal matters, including with issues such as career or marital problems.[27] This type of support is missing from the Northern Irish system.

Initial research undertaken by the authors on wellbeing across the legal profession in the UK and Republic of Ireland indicated that these distinct features could themselves potentially have an impact upon the wellbeing of barristers in Northern Ireland.[28] The 2015 Wellbeing at the Bar report for the Bar Council of England & Wales highlighted the level of support from others within chambers as a positive factor in protecting and enhancing wellbeing.[29] Sixty-six per cent of respondents found that 'Peers and clerks were sources of support most or all the time.'[30] Therefore, the lack of these positive resources at the Northern Ireland Bar is potentially significant. Legal professionals also portrayed the Bar in Northern Ireland as a 'highly competitive environment, with a glut of qualified and working barristers'.[31]

25 Roger Bowles, 'The structure of the legal profession in England and Wales' (1994) 10 Oxford Review of Economic Policy 18.

26 John Flood, "He's fucking marvellous!': the fall and rise of barristers' clerks' (2007) in *The Fall and Rise of Barristers' Clerks* (not published in a journal) 5.

27 John A Flood, *Barristers' Clerks: The Law's Middlemen* (Manchester University Press 1983) 55.

28 Jones et al (n 8 above) 78.

29 Positive (n 1 above).

30 Ibid 9.

31 Ibid.

AIMS AND METHODOLOGY

This study aims to fill a gap in the literature by assessing perceptions of wellbeing[32] at the Northern Ireland Bar. Based on 10 semi-structured interviews and adopting an inductive thematic analysis of the data, this article will argue that working at the Northern Ireland Bar places a range of demands on individuals, potentially decreasing their engagement with their work and adversely impacting their wellbeing. These include difficulties in becoming established and developing a legal career, work–life balance issues, pay concerns, inter-relational demands and structural inequalities. However, in their work barristers also have positive experiences and motivators beneficial to their wellbeing, including being able to engage in interesting and sometimes enjoyable work, collegiate working practices, and emerging initiatives aimed at better supporting practitioners.

In identifying demands and resources, the findings of this study correlate well with the approach of the job demands and resources (JD-R) model.[33] This is a model for evaluating workplace wellbeing through the identification of job demands and job resources. Job demands are those 'physical, psychological, social, or organisational aspects of the job that require sustained physical and/or psychological (cognitive and emotional) effort or skills'.[34] Such demands have physiological and psychological costs which can lead to burnout and exhaustion, as well as a lack of engagement and motivation.[35] In contrast, job resources are those aspects of work which are either 'functional in achieving work goals', 'stimulate personal growth, learning, and development' or 'reduce job demands and the associated physiological and psychological

32 The definition of wellbeing adopted in this study (and provided to participants in the Project Information Sheet) is that of the World Health Organization which refers to optimal psychological wellbeing as 'every individual realizes his or her own potential, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to her or his community': *Promoting Mental Health: Concepts, Emerging Evidence, Practice* (Summary Report) (World Health Organization 2004).

33 Evangelia Demerouti et al, 'The job demands-resources model of burnout' (2001) 86 Journal of Applied Psychology 499; Arnold B Bakker and Evangelia Demerouti, 'The job demands–resources model: state of the art' (2007) 22 Journal of Managerial Psychology 309.

34 Evangelia Demerouti and Arnold B Bakker, 'The job demands-resources model: challenges for future research' (2011) 37 South African Journal of Industrial Psychology 1, 2.

35 Demerouti et al (n 33) 502. Crawford et al distinguish between 'challenge demands' and 'hindrance demands', suggesting it is the latter which impact negatively on engagement and motivation: E R Crawford, J A LePine and B L Rich, 'Linking job demands and resources to employee engagement and burnout: a theoretical extension and meta-analytic test' 95(5) Journal of Applied Psychology 834–848, 843.

costs'.[36] The latter of these demonstrates that job resources can have a 'buffer' effect against high levels of job demands, ameliorating the potential consequences of these.[37] Overall, job resources will increase engagement and promote positive wellbeing and enhanced commitment and motivation.[38] The JD-R model is commonly applied within professional contexts, including teaching and the police.[39] It has also increasingly been applied to the legal profession in a range of jurisdictions, including the UK, Australia and South Africa.[40] In a longitudinal study of the model, it was found to provide a 'valuable heuristic tool' for promoting wellbeing at work.[41] In this study, it was noted that one limitation of this model was its tendency to focus upon workplace factors ,whereas it was found that 'home demands and home resources' were also factors which should be considered.[42] This is also illustrated by the themes that emerged within the current paper, in particular, that of 'work–life balance'. A further longitudinal study also supported the JD-R model overall, but noted that job resources may not 'buffer' job demands to the extent originally suggested, meaning a greater focus on reducing job demands may be required.[43] Drawing upon this model, this paper will suggest that key stakeholders in the Northern Ireland Bar, such as the Bar Library, should seek to reduce

36 Bakker and Demerouti (n 33 above) 2.

37 Arnold B Bakker, Evangelia Demerouti and Martin C Euwema, 'Job resources buffer the impact of job demands on burnout' (2005) 10 Journal of Occupational Health Psychology 170; Bakker and Demerouti (n 33 above) 317; Bergin and Jimmieson (n 2 above) 438.

38 Arnold Bakker, Evangelia Demerouti and Wilmar Schaufeli, 'Dual processes at work in a call centre: an application of the job demands-resources model' (2003) 12 European Journal of Work and Organizational Psychology 4, 393.

39 See, for example, T Dicke, F Stebner, C Linninger, M Kunter and D Leutner, 'A longitudinal study of teachers' occupational well-being: applying the job demands-resources model' (2018) 23(2) Journal of Occupational Health Psychology 262. Dicke et al note that the JD-R Model is 'well established in occupational research'. Also note, R Scheepers, M Silkens, J van den Berg et al, 'Associations between job demands, job resources and patient-related burnout among physicians: results from a multicentre observational study' (2020) BMJ Open 10:e03846. Scheepers et al describe the model as 'validated' (at 2).

40 Jones et al (n 8 above) chapter 5; Bergin and Jimmieson (n 2 above); Elsie Rossouw, and Sebastiaan Rothmann, 'Job demands and job resources and well-being of judges in South Africa' (2020) 46 South African Journal of Industrial Psychology/SA Tydskrif vir Bedryfsielkunde 00, a1801.

41 Jari J Hakanen, Wilmar B Schaufeli and Kirsi Ahola, 'The job demands resources model: a three-year cross-lagged study of burnout, depression, commitment, and work engagement' (2008) 22(3) Work and Stress 224.

42 Ibid 235.

43 Anna-Carin Fagerlind, Chrisian Ståhl and Peter Smith, 'Longitudinal association between psychological demands and burnout for employees experiencing a high versus a low degree of job resources' (2018) 18(1) BMC Public Health 915.

the levels of job demands upon barristers where possible. Where this is not possible, for example, because of the broad structural nature of the issues, the focus should be on maximising the resources available to barristers to ameliorate, at least to some extent, the potentially detrimental impacts on wellbeing of high levels of job demands.

The sample for this study consisted of all barristers currently eligible to practise in Northern Ireland. A snowball sampling technique was used to obtain participants, initially via invitations to existing contacts who then introduced additional potential invitees. Four participants had experience of between 1 to 5 years of call (experience since qualification) at the Bar Library, two participants had 5 to 10 years of experience, two participants had 10 to 15 years of experience, one participant had 15 to 20 years of experience, with one barrister with 35 to 40 years of experience. All participants, except for one, were enrolled with the Bar Library: this participant had stopped practising for the Northern Ireland Bar, but retained a practising certificate in their current employment. Two of the participants had previous legal experience – one participant had worked as a paralegal for one year, while another had experience working in the charitable sector in a legal role. Some of the participants disclosed working jobs additional to their barrister's role, not all of which were legal related. The participants had experience in a broad range of practice areas, including commercial law, children's order work, family law, criminal law, employment law, housing law, human rights, and immigration law.

All the interviews were conducted face-to-face in June and July 2019. Participants were invited to reflect on their perceptions of wellbeing. They were asked to discuss whether levels of emotional wellbeing at the Bar have changed, the key factors which influence wellbeing, and whether they have experienced hostile behaviours from clients, peers, or judges. They were also invited to raise other issues which might directly impact on their wellbeing. Questions included:

- How would you describe the emotional wellbeing of those working in your sector of the legal profession?
- Do you find working as a self-employed barrister challenging? If so, why?
- Do you perceive inequalities in the profession?
- Have you always been motivated to work as a barrister?
- Do you think you can rely on other barristers for support?
- Do you think working in chambers or joining a trade union would improve your working conditions?

A semi-structured approach to interviewing was adopted to allow flexibility and comparability. It also allowed further investigation of

points made by the interviewees to gain a more holistic view of their perceptions.

An inductive thematic analysis was adopted to analysing transcripts of the interviews, allowing for rich and compelling insights.[44] Both researchers jointly coded one interview, before individually coding the remaining transcripts. All coding was reviewed by both researchers. The researchers then jointly formulated themes and sub-themes from the coding using the qualitative data analysis software NVivo12. The key themes include:

- establishing and developing a legal practice;
- work–life balance;
- financial issues;
- professional relationships;
- inequalities at the Bar; and
- the practices and cultures of the Bar Library.

The use of a snowball technique to obtain participants can be viewed as a limitation of this study, given its status as a form of 'convenience sampling' rather than the arguably more representative sample obtained via more purposive techniques.[45] However, its value in obtaining participants from within 'hard-to-reach' populations, including elites, is well established.[46] Given the small, close-knit, nature of the legal community in Northern Ireland, and the self-employed status of the participants, it was therefore viewed as appropriate. The relatively small number of participants could also be viewed as a potential limitation. However, participants with various levels of experience and differing practice areas were included within the sample, all of whom provided rich and insightful data, allowing for the emergence of important themes and, consequently, an important contribution to work in this area.

44 Virginia Braun and Victoria Clarke, 'Using thematic analysis in psychology' (2006) 3 Qualitative Research in Psychology 77; Virginia Braun and Victoria Clarke, 'What can "thematic analysis" offer health and wellbeing researchers?' (2014) 9 International Journal of Qualitative Studies on Health and Well-Being 1; Alan Bryman, *Social Research Methods* (Oxford University Press 2016).

45 Mark S Handcock and Krista J Gile, 'Comment: on the concept of snowball sampling' (2011) Sociological Methodology 367, 368.

46 Rowland Atkinson and John Flint, 'Accessing hidden and hard-to-reach populations: snowball research strategies' (2001) 33 Social Research Update 1, 1.

ESTABLISHING AND DEVELOPING A LEGAL PRACTICE

Lack of preparedness

Participants in this study indicated that getting established and making an economically viable career from their work was one of the biggest challenges facing barristers in Northern Ireland. The early stages in the life of a barrister are crucial in terms of being able to establish a sustainable long-term career, yet barristers stated that they considered themselves unprepared for life as a barrister, found themselves shocked at the reality of the experience, and struggled to forge a career within the profession.

One of the issues raised by several participants was a lack of preparation for the practical realities of practising as a barrister. The limitations of the academic and professional stages of qualification were particularly highlighted, with one participant reflecting that their professional training had prepared them well for advocacy, but not for other aspects of the role:

> We never were told how to write an attendance note or about doing your fees, chasing your fees, dealing with legal aid, the real practicalities of things of how you actually survive those early times … You were just looking for tips from other people, how do you do this? I think that was the main thing that was probably lacking, which really could support you in the early days. (Barrister 9)

Other participants raised similar points, highlighting ethics, accounting and bookkeeping, practice management, marketing and branding, and wellbeing, as topics with insufficient coverage. Some participants suggested that it would be difficult to incorporate an awareness of the issues into these stages, suggesting it was 'just the kind of thing you have to learn as you go' (Barrister 5). However, one participant who had taken a clinical legal education module during their undergraduate degree highlighted the value this could have in introducing students to real-life issues.

The experience of pupillage itself (the one year of on-the-job training, shadowing a pupil supervisor)[47] was also touched on by participants, with some viewing it as providing valuable experiential learning, but others suggesting that there remained a lack of tangible experience because 'you're not getting it full-on' (Barrister 2). Several participants referred to an absence of structure during pupillages, which could lead to a lack of feedback and skills progression, contributing to a sense of

47 A pupil supervisor is referred to as a pupil master in Northern Ireland. A female master has sometimes been referred to in England and Wales as a pupil mistress, but this does not appear to be a term which is commonly used in Northern Ireland where pupil master covers both genders.

unpreparedness. One participant attributed this to the fact that 'anyone over seven years call can be a pupil master', with the 'checklists' in place to structure the pupillage 'not really being followed' (Barrister 1). For those who had found it a positive experience, the emphasis seemed to be on the supportive personal relationships developed with, and through, the individual master in question, rather than on the experiential learning involved.

Practical demands

Alongside their lack of preparedness, the practical demands of the role were also emphasised by participants. One participant depicted the demands placed on them, creating multiple roles which a barrister performs:

> [T]hat's you, self-employed. And it takes you a long time to realise that you are a business, a one-man band business. And it's not just being an advocate and going to court. You have to be a secretary, to do your paperwork; you have to be a receptionist, to take your phone calls; you have to be admin ... you have to be like a social worker, you have to be a priest, you have to listen, you have to ... . (Barrister 2)

This reference to the self-employed nature of the role and lack of administrative support appears to, at least in part, reflect the distinct nature of the Northern Ireland Bar. In contrast, as stated above, in England, Wales and Scotland it is usual for chambers and stables to have barrister's clerks, who are responsible for agreeing fees on behalf of barristers and scheduling work, relieving them of some of the administrative load.

Lack of job security

Another issue raised, which again relates to the self-employed nature of the role, is the lack of job security. The Northern Ireland Bar is a 'referral bar' meaning that the usual route for a barrister is to be instructed through a solicitor. Unlike in chambers or stables, there are no clerks acting as intermediaries – the relationship between barristers and solicitors is much more direct. The relationship that barristers have with practising solicitors is therefore crucial, but can be slow to develop with no roadmap for new entrants into the profession. This is one barrister's experience of the early stage of their career and the shock of understanding that, upon becoming a barrister, they were not secure within their profession:

> It was a shock ... you've gone through your law degree, you've done Institute exams, you've done the Institute, the whole way through you're going to be a barrister coming out of this, automatically your head goes to, secure for life, money and all this and status ... . There's the status things and the money can be there month to month, but we're

> not told about the real realities of [...] you can have a great month, but then you might have three months with no money coming in. You'll have months of just picking up scraps ... just trying to build a practice. There'd be weeks where you might not have a case. How you deal with that definitely wasn't portrayed. You have people coming to you all the time and saying, 'look, it's not as easy, it's not as glamorous'. It was never really painted like that. (Barrister 9)

Participants described how at the start they had to 'take everything, no matter what comes your way'. Several shared stories of having to go to the Bar Library in the hope that some work would become available, for example, if another barrister needed a hearing covered. However, several participants also described how they would turn up to the Bar Library and find that they had nothing to do:

> We were coming in half seven, eight o'clock, to sit down there for the morning. Some mornings it's just nothing. You'd be sitting there watching TV shows on your laptop, you literally had no work to do. Somebody would come down from the reception and say, 'there's a case, can somebody cover it?' It's whoever puts their hand up first. (Barrister 9)

Progression at the Bar

Individuals' progress within the Bar was described by the participants as a very uncertain process, involving part luck and, possibly part, nepotism. Unlike other roles within the legal progression, as one participant remarked, 'there is no set progression ... no career path as such'. One practitioner who had retired from the Northern Ireland Bar cited the lack of career progression as the reason for them leaving. This participant indicated that there was an unfairness in the system.

The Bar in Northern Ireland does not have statistics on how many enrol within a year and how many leave the Bar. However, participants reported seeing many of their peers leaving, often for financial reasons. The effect of individuals having to quit the Bar was discussed by some of the participants, who suggested it was associated with feelings of failure. Having to leave the Bar may be particularly emotionally difficult for some barristers to countenance, as becoming a barrister is a select profession which requires a long process of education and training. One of the participants spoke of a conversation she had with a friend within the Bar who was considering leaving, who explained they were feeling like a 'failure' and questioned: 'What will my peers think of me? What will my family think of me?' (Barrister 3)

At the same time, there was a consensus that at some point, if a barrister managed to stay in their role, they might become established in their careers, financially their practice would become viable, and they could start earning large sums of money. The first five years of

practice were mentioned as being particularly important to barristers. Not only is this because it is the early part of their career, but within the first seven years barristers are able to pay a reduced rate annually to the Bar Library for their services. Some participants reflected that, after seven years, barristers should have a sustainable practice operating. Therefore, there is a pressure on barristers within the first years of their practice to take work on regardless of the circumstances. It also means that there is pressure on barristers to perform in the cases that they are working on.

> Every case that we do, especially in the first couple of years, seems to be massively important, because we don't know whether we're going to annoy or please our solicitor, we don't know, even in terms of writing letters to them, whether we're doing in the right form or not. (Barrister 7)

Participants suggested that they may take on too much work and experience poor work–life balance as a result of this commitment. As one interviewee commented, 'people do tend to kill themselves with just taking on as much as they can, whenever it's available' (Barrister 10).

Overall, this theme emphasises the high level of job demands placed on junior barristers as they seek to establish a financially viable career while navigating heavy practical demands. Such demands appear to be largely structural in nature, often generated by the self-employed nature of the profession and its competitive quality, characteristics which are not unique to Northern Ireland, but appear to be ubiquitous across jurisdictions in relation to the self-employed Bar.[48] However, applying the JD-R model, it is possible that these job demands are exacerbated by the specific lack of administrative support at the Northern Ireland Bar. There do appear to be opportunities to ameliorate these demands through an increase in relevant job resources. This could be, first, through a review of the vocational training and pupillage system, to ensure that they manage expectations and focus on equipping aspiring barristers with appropriate skills and coping mechanisms for dealing with structural issues. Second, through a consideration of whether additional administrative support can be provided via the Bar Library to assist in alleviating some of the practical job demands.

48 Jones et al (n 8 above) chapter 2; Positive (n 1 above) 24.

WORK–LIFE BALANCE

The term 'work–life balance' is commonly used but also contested and culturally and temporally situated.[49] In relation to the legal profession, it has been argued that the concept is problematic as the use of flexible working and technological advances can lead to a 'work/work' culture, dissolving work–life boundaries.[50] However, for the purposes of this theme it is used as a useful shorthand to refer to 'the absence of work–life conflict – that is, work should not consume all of one's energies and there should be time for family life, socialising, rest and relaxation, as well as time to pursue interests in music, theatre, art or sport'.[51]

Work levels

The perceived necessity of taking on all available work discussed above was not exclusive to junior barristers, it was also shared by a number of the more senior barristers interviewed. Several participants reflected on how they might feel they are achieving a position where they are getting steady work over a stretch of time, only for that to be followed by a period where they have less work on. Although there was a recognition that some who are particularly well established within the profession can slow down their pace of work, the uncertainty around the availability of work appears to mean that many barristers take on too much work, affecting their work–life balance and ultimately their wellbeing.

> [T]here's always a constant thing because … they call it feast and famine, but because you will suddenly get, even when you're fairly busy, got a good reputation […] a sudden rush of work then it may dry up for a while. So, that's a constant, no matter how busy you are or not, that's a constant temptation to just take on work. Even if you're double booking, triple booking yourself, you take on work because it may not be there in a given month. (Barrister 8)

The widespread perceived need to continually take on work to make a practice a profitable enterprise appears to mean that, even when barristers are busy, they will still feel the need to take on further work. It was suggested that continually working, including on holidays and during unsociable hours, is common practice amongst most barristers. This appears to be partly to do with a fear of working drying up,

---

49 See, for example, Suzan Lewis, Richenda Gambles and Rhona Rapoport, 'The constraints of a "work–life balance" approach: an international perspective' (2007) 18(3) International Journal of Human Resource Management 360–373.

50 Thornton (n 5 above) 35.

51 Margaret Thornton, 'The flexible cyborg: work–life balance in legal practice' (2016) 38 Sydney Law Review 1, 8.

although for others it may represent an ingrained habit, or a focus on increased financial rewards.

The participant below reflects that this has to do with the competition at the Bar, which places high demands on individuals – if you are not ready to do the work then solicitors may instruct someone else who can.

> The demands on the profession are still quite high and there's still a culture of people emailing you late at night, early in the morning, expecting you to be ready to respond to that. And the other thing that I find is even when you're on holiday you're expected to be contactable and you're expected to be able to do things for people if they need them done. Because there is such competition at the bar that if you don't do those things, someone else will ... and solicitors move on. (Barrister 4)

Views on whether the expectation to be contactable on holidays is a profession-wide phenomenon varied among participants. One female barrister reflected that it may be a gendered practice to feel unable to take breaks, noting that a male colleague explained they could go on holidays abroad each year without their laptop. However, some male participants in the study indicated that they needed to be contactable when taking holidays. In a general sense, many of the participants felt unable to take proper restful breaks from work:

> [I]t's very difficult to take breaks. It's always on to the next thing. You go into court, and you come out of court in the morning, maybe it was a great result or whatever. And the next, you're just straight back into it. (Barrister 1)

It was clear from the interviews that barristers often work long and unsociable hours. For example, some of the barristers talked about sending and receiving emails at three o'clock in the morning to and from other barristers or instructing solicitors. Another participant stated:

> [There are] intense periods of stress when a lot of the time we get stuff very last minute. So, I could be getting papers at ten o'clock, 11 o'clock in the evening emailed across to me and I'm expected to turn that around for the next morning. (Barrister 4)

Although many participants spoke negatively of the inability to take breaks from work, others spoke positively of the flexibility of being a self-employed barrister:

> I like being my own boss. I like being able to manage my own time, albeit I'm not very good at it. And it does lead to working at all hours of the night. But it means that if I want to take time off, I can take time off as well. (Barrister 10)

Other barristers discussed the effect of technology creep on their working lives, indicating that the use of telephone and email communications

had eroded work–life boundaries. This is not only the case with regards to working in the evening, but when barristers are on their holidays.

> The needing to be constantly available, the emails coming to your phone, and always on it, and there's no line ...There's no line between working and not working. You're always on it. (Barrister 3)

One of the barristers reflected on that fact that she had developed mechanisms to control her work–life balance, but gradually over time she had reverted to what she was previously doing. This indicates a danger in that, even if individuals are conscious of the detrimental effects of having a poor work–life balance and take steps to address it, work still has the capacity to take a more central role in a person's life without continued effort to ensure that this does not happen.

Impacts of current working practices

Several of the participants referred to constantly ruminating on their work, even during times when they were not working, which was discussed by some as an inability to 'switch off'.

> It's very difficult to switch off from cases. Especially if they're cases where I would say it's for a vulnerable person or it's a case that they care deeply about. And you feel the burden of their distress over it. It's all very well and one could say, 'oh the court doesn't sit in July and August and over Easter'. But what I'm really talking about is that that break doesn't reflect a break from the way of thinking about the problems that you've got. (Barrister 1)

Many barristers found it hard to rest and relax, even when not working, meaning that they had no time to recover from the stresses and demands of work. Such an inability to recover can manifest itself, for example, in insomnia and subjective sleep complaints,[52] which have an obvious connection to wellbeing. For those individuals with perfectionist tendencies, like many within the legal profession,[53] this inability to switch off is likely to be heightened.[54]

As well as the impact upon individuals, there was also a discussion regarding the impact of work levels on family life. One of the participants saw the in-built flexibility of being a barrister as family-friendly, although recognised that it is always tempting to take too much work on. This idea of the Bar being family-friendly was also recognised by a female advocate in Melville and Stephen's study of the Faculty of

52 Margareta Kristenson, Hege R Eriksen, Judith K Sluiter, Dagmar Starke and H Ursin, 'Psychobiological mechanisms of socioeconomic differences in health' (2004) 58(8) Social Science and Medicine 1511.

53 Jones et al (n 8 above).

54 Cecilie Schou Andreasse, Jørn Hetland, Helge Molde and Ståle Pallesen, '"Workaholism" and potential outcomes in well-being and health in a cross-occupational sample' (2011) 27(3) Stress and Health 209.

Advocates in Scotland.[55] Other barristers reflected on the adverse impacts that their working schedule had on members of their family:

> [Y]ou have to give it your all. So, whenever something happens at home where, say, your husband takes sick, your child takes sick, it's really difficult … Or, even, you take sick yourself, to keep […] all the balls in the air, you know. You know, because your family life does suffer. (Barrister 3)

Long working hours and heavy demands upon time are a well-established theme within the wider literature on lawyer wellbeing, including within a previous application of the JD-R model and amongst barristers in the Australian context.[56] Although barristers are not subject to chargeable hours and billing targets, it is clear that the self-employed competitive nature of the Bar requires lengthy unsociable hours, often exacerbated through technology creep. The impacts of this in terms of work–life balance and wellbeing are significant. The comment above also demonstrates that the workplace cannot be viewed as a wholly discrete domain, rather there is a need to consider wider demands and resources that impact upon, or are impacted by, practices at the Bar.

FINANCIAL ISSUES

As discussed above, for many barristers, particularly junior practitioners, it can be difficult to develop a financially sustainable practice. A central issue affecting barristers is being paid late, being paid small amounts, or not being paid at all. This disproportionally affects junior barristers because they are unlikely to work on as many high-earning cases. They may also work for free on occasion, to establish relationships with a solicitor's firms and raise their professional profile. This can lead to exploitation, as one participant explained:

> Doing work that I don't get paid for, there's two types of solicitors that I've had that with, there've been solicitors who've been really upfront with me and said, look, I'm really sorry about this, I have to do this as a favour to my client, I wonder if you'll do as a favour to me? There are these other cases that will come your way, I do see more work from them, and they've developed into really good relationship with the solicitors. That I see as a loss leader, and I'm happy to do it. I'm happy to build that relationship with solicitors. Other solicitors just don't pay

55 Angela L Melville and Frank H Stephen, 'The more things change, the more they stay the same: explaining stratification within the Faculty of Advocates, Scotland' (2011) 18 International Journal of the Legal Profession 211, 217.

56 Bergin and Jimmeson (n 2 above); Bernadette Healy, 'Towards a relational perspective – a practical and practice-based discussion on health and wellbeing amongst a sample of barristers' (2014) 14 Queensland University of Technology Law Review 94.

> me and take advantage of the fact that they know I'm young and that's incredibly frustrating. (Barrister 7)

Another participant reflected on the fact that, while they might get paid for work, they were getting paid small amounts of money and were having to travel significant distances for it. In this case, the participant is discussing travelling to Derry which is 1 hour 40 minutes by car from Belfast (70 miles):

> I could get a mention up in Derry … I'd get the bus up there. But it could be, like, a £10 or £15 mention and you may never get paid. (Barrister 8)

Several of the barristers, particularly those who were early career, discussed having part-time jobs to supplement the wages that they were getting from working at the Bar. Some of those jobs might have been legally related, although others were not.

> I was going to say, it's been relatively okay … But every year, when I do my accounts, I look back over my whole career and what I haven't been paid. And when I did that in January, I'm owed £30,000, which is a lot of money that I could do a lot with … The impact of not getting money in is that you then have to support yourself somehow else. And a lot of people have part time jobs for the first few years at least, until they have a steady stream of income. And in fact, I would suggest that probably most people, if not everybody, has some form of alternative stream of income, unless they're really wealthy, or their parents are really wealthy, and able to keep them. (Barrister 10)

At the same time, it was acknowledged that being at the Bar could bring significant financial benefits, particularly once established:

> You see, we're lucky in a way. Our legal aid system is about ten to 15 years behind England and Wales. We haven't had LASPO[57] or anything like that, so we're still relatively well remunerated. (Barrister 4)

It is junior barristers who are particularly vulnerable to challenges with financial issues, making it a potentially significant job demand. Applying the JD-R model, whilst the promise of future financial stability could be a motivational job resource, it is questionable to what extent it is enough, given the additional pressures upon early practitioners discussed in the section above on 'Establishing and developing a legal practice'.

57 Legal Aid, Sentencing and Punishment of Offenders Act 2012.

PROFESSIONAL RELATIONSHIPS

Collegiality

The Bar Library seeks to encourage a culture of collegiality amongst its members to aid in the administration of justice. In a submission from the Bar Council in Northern Ireland to the Committee of Justice of the Northern Ireland Assembly, the Bar Library system and its culture of collegiality is explained, with specific reference to its benefits with respect to the conflict in Northern Ireland:

> The consequences of all barristers working together from the same building using the same facilities and sharing the same ethos is that the religious and political differences that have so disfigured Northern Ireland have not been permitted to operate. It has facilitated unhindered access to legal representation for many unpopular causes throughout the troubled history of Northern Ireland. The cohesion and collegiality of the Bar has thus ensured a broad acceptance of the impartiality of the Northern Ireland legal system, thus aiding the administration of justice, a fact which has been acknowledged on many occasions by the judiciary and successive governments.[58]

The Bar Library operates a 'family system' which begins from the appointment of a pupil master to a barrister during his or her pupillage. The duties and responsibilities of the pupil master are laid out within the Code of Conduct of the Bar of Northern Ireland; however, the emergence of the pupil master's family appears never to have been written down and is a cultural, rather than regulated, facet of the Bar. The following participant describes the structure of the Bar family and how it creates collegiality within the Bar Library environment, including connections to your pupil master's previous and subsequent pupils (brothers and sisters), and their master's own pupil masters (grandparents):

> [You have] brothers and sisters, and [...] even grandparents. You'll find people will go back to their master's master and getting help about things. But having that network, that's the way that I think people get initiated into the collegiality. Because, through my Bar sisters who are younger than me, I know a huge number of other people in their year. And they come to me for help sometimes. (Barrister 10)

This comment reflects the experience of several of the participants – that a barrister's master, or their Bar family, is a place to turn to for support. Through these relationships, barristers can build wider relationships with other members of the profession. However, views among participants on the extent of collegiality varied. Many of the

58 Submission from The Bar Council to the Committee for Justice (2009).

participants were very positive about their experience of collegiate ways of working. For example, one participant stated:

> I think it is really a collegiate place [...] I think albeit everybody is self-employed, and what is my brief one day could be somebody else's another day, people will not hesitate to help each other out. (Barrister 10)

However, the notion of collegiality was challenged by several of the participants who perceived the competition at the Bar as eroding this:

> [T]here's certain people you go to. You wouldn't go to your direct competitors. You're all competing, at the end of the day. It's cut-throat, in terms of keeping cases. If you see somebody as a threat, you're never going to recommend them to your solicitor to cover something. (Barrister 7)

> The plaster that was put on it for a long time and still is thrown about, is this collegiate type notion. That barristers are a part of this Library system and the Library system and the Bar is this collegiate body of people that look after and support each other. That is not my experience and it's not the experience of others. It is a plaster that's put on to a problem. (Barrister 1)

Negative interactions with colleagues and judges

Hostility between colleagues, or from other members of the legal profession, was identified by participants as having the potential to adversely affect their wellbeing. Participants discussed a range of negative behaviours between individuals, ranging from discourteousness, to what might be described as bullying. Some participants reflected on the fact that the adversarial system has a bearing on how individuals within the profession treat one another, with one participant characterising confrontation as a cultural facet of this system.

> Well, it can be reflected in just the day to day dealings that you have on the other side, because, it's an adversarial system. So, you do have run-ins with practitioners on the other side. (Barrister 8)

Other participants discussed the 'style' of some barristers as being aggressive.

> [There] would be certain people who I would be very reluctant to be on the other side of because of the way that they act. It's nothing against me. It's just the way that they act towards every single person. It's just their style. Certain people have a very aggressive style ... .(Barrister 5)

One participant referred to forms of 'territory marking' where barristers were deliberately being discourteous to prove their dominance or

seniority, presumably to project a persona of professional superiority and attempt to destroy their adversary's confidence. These attitudes again appear to reflect the adversarial nature of the Bar:

> [T]he very prospect of actually discussing a case with them, fills one with anxiety and distress because you know where it's going to end up with some sort of personal slight or nasty comment. It's not, '[my case has] value and is worth this because of this case. Then they respond back saying, well our point is this and we consider that, and we say your point's not good because of this logical reason'. We get [...] 'well that's a load of crap' and then they walk off [...] there's this sort of territory marking thing where 'I'm older than you [...] I'm better. I've been at this a long time and your case is the worst case I've ever seen. This is disgraceful'. (Barrister 1)

As well as the impact upon wellbeing of interactions with other barristers, a common theme in interviews was the impact of interactions with members of the judiciary. Whilst it was recognised that judges challenging barristers on their knowledge of the applicable law, aspects of their cross-examinations and their use of legal arguments had a beneficial function (enhancing court proceedings and ultimately access to justice), a more pernicious form of challenge was described by some participants:

> I've seen a judge make a fool out of somebody, deliberately set out to embarrass them in court. That would be devastating for your self-esteem and your self-confidence. (Barrister 5).

Several participants highlighted that younger members of the profession were more frequently challenged in a more disrespectful manner, with one participant remarking:

> [Y]ou have some other judges who, I'll maybe not describe in any more detail, but who enjoy putting new faces through their paces. There's one thing to do that as a [...] learning practice, but my experience is that those who do that, do it just out of pleasure or enjoyment of seeing new people squirm. That can really put people off, especially in a full court room. I've developed a fairly thick skin and now something like that wouldn't bother me, but certainly in the first year out, to be publicly criticised in a room full of 20 of your colleagues who are all more senior makes you question why you're doing the job in the first place. (Barrister 7)

One participant, an early career barrister, described how a judge who could 'be quite sexist towards young female counsel' had cut her off in the middle of a cross-examination and 'shot her down' on every point she had been trying to make to a witness. Describing how she felt afterwards, she stated:

> I was in tears, and I was like, that was the worst day, you know, my worst day at the Bar. Because I just ... I felt that I'd put a lot of effort into the case, I felt humiliated in front of the client, and obviously I

> had to hold it together, you know, for them, and just say 'oh, you know, that's the way that judge is, you know, that's his style'... But I had a lot of justifying to the client and I just kind of felt if he had a male barrister representing him, he wouldn't have been treated in that way. (Barrister 3)

None of the participants suggested that the Northern Ireland judiciary in its entirety was hostile or negative towards barristers. Instead, it was indicated that there were a minority of specific individuals whose behaviours were inappropriate. As one participant noted, 'They're not all like that, it's a small handful ... the vast majority are very understanding, of course, they just want respect and people to be well prepared.' (Barrister 9). Nonetheless, as several participants recognised, the behaviours of this minority can be detrimental to wellbeing, especially for early career barristers who are likely to be dealing with the myriad of other issues around establishing their career.

The Positive Report on the Bar of England and Wales noted that 73 per cent of respondents 'endorsed that there was a sense of cooperation and collaboration in their work environment most or all the time' with 66 per cent referring positively to peers and clerks.[59] Therefore, outside of the Northern Ireland jurisdiction there is a sense of collegiality amongst colleagues, but the fact that a sizeable minority did not endorse the above statement corroborates our findings that cooperation and collaboration can be lacking. Collegiality within the workplace has a direct relationship to wellbeing.[60] For example, in a study of legal academics in the United States by Seigel and Miner-Rubino, it was found that uncollegial behaviour can be harmful to the wellbeing of individual faculty members,[61] while there was a moderate positive correlation between collegiality and job satisfaction.[62] Conversely, incivility in the workplace has a negative effect on wellbeing. Pearson and Porath, for example, surveyed 700 employees in different occupational settings and found that as employees' experiences of incivility increased, their level of job satisfaction decreased.[63] Applying the JD-R model, it is clear that the notion of a Bar family is a significant job resource to barristers in Northern Ireland, one which supports wellbeing through fostering collegiality. More widely, it may be that the promotion of collegiality

59 Positive (n 1 above) 9.

60 Jones et al (n 8 above) chapter 3.

61 Michael L Seigel and Kathi Miner-Rubino, 'Measuring the value of collegiality among law professors' (2009) 1 Faulkner Law Review 257, 280.

62 Ibid 279.

63 Christine M Pearson and Christine L Porath, 'On the nature, consequences and remedies of workplace incivility: no time for "nice"? Think again' (2005) 19 Academy of Management Perspectives 7, 8.

also goes some way towards preventing the development of entrenched competition between members. However, again applying the JD-R model, it can be noted that the incivility identified in this study suggests that there remain job demands when dealing with both colleagues and the judiciary. Once again, this appears to manifest itself most clearly in relation to junior barristers, where there appears to be a quasi-form of 'initiation ceremony' involving being challenged by judges, probably in order to test their credentials. This requires careful consideration to be given to ways in which resources such as the Bar family concept can be enhanced to buffer such demands as effectively as possible.

INEQUALITIES AT THE BAR

Gender inequalities

Disability, age and gender, were all discussed as being areas of inequality at the Northern Ireland Bar. However, the dominant issue concerned gender inequality. Gender differences and gender inequality are entrenched at the Bar in other jurisdictions within the UK.[64] The Bar has traditionally been a male-dominated profession.[65] In the early twentieth century women were not allowed to join the legal profession, and it was not until 1919 that Helena Normanton became the first woman to join the Middle Temple in England following the introduction of the Sex Disqualification (Removal) Act 1919.[66] In Ireland, Frances Kyle was the first female to be called to the Bar in November 1921, however, Averil Deverill was the first woman to practise as a barrister in Ireland (this was before the partition of Ireland).[67] The next barrister to follow her, in Northern Ireland, was Sheelagh Murnaghan who was called in 1947, almost 30 years later.[68] Former President of Ireland, Mary McAleese, recalls that when she started studying law in 1969 there were no women practising at the Bar in Northern Ireland and only about 20 per cent of the law class was female.[69]

64 Suzanne McKie and Ruth Whittaker, *Statistics and Analyses Regarding the Slow Progression of Women in Professional Spheres* (Farore Law 2019) 11; The Bar Council, *Momentum Measures: Creating a Diverse Profession* (The Bar Council 2015).

65 Patrick Polden, 'Portia's progress: women at the Bar in England, 1919–1939' (2005) 12 International Journal of the Legal Profession 293.

66 Ibid.

67 Erika Rackley and Rosemary Auchmuty, *Women's Legal Landmarks: Celebrating the History of Women and Law in the UK and Ireland* (Bloomsbury Publishing 2018).

68 Mary McAleese, *Celebrating a Centenary of Women in Law* (Celebrating a Century 2019) 3.

69 Ibid.

Although the Bar is no longer such a male-dominated environment, the legacy of patriarchy in the Bar remains, with many of the participants reporting several gendering practices. In addition to specific issues, the Bar requires individuals to work unsociable hours, which can directly have an impact on family caring and childcare responsibilities, for which – as Collier argues – women carry most of the burden.[70] The culture of the legal profession, as argued by Collier and Sommerlad, is hyper-competitive and hyper-masculine.[71] Separate research by the authors discusses that this leads to an emphasis within legal culture on perceived 'masculine' conceptions of strength, the ability to continue work despite challenges, and work taking precedence over illness or emotional issues.[72]

The need to show strength was something discussed by the participants in the study. This was explained as being linked to being able to run a financially viable practice and also to the perception that if solicitors thought a barrister could not perform, they would lose their business:

> [Y]ou don't want to be showing a sign of weakness. [...] it's not just your colleagues at the bar. It also would be solicitors. You wouldn't want to show any sign of weakness, for fear of not getting instructed. (Barrister 2)

Another participant corroborated this, stating:

> That's the problem with being a barrister. You have to portray a very confident, assertive outlook. You have to act the part. If you drop that mask, or if you seem to be weak or not well, then the concern is that you might lose your practice. You might lose your trusted solicitors and things like that. (Barrister 5)

The same practitioner also reflected that it could be difficult to seek help from people, for fear of appearing weak: 'With your going and looking for help from someone, you're exposing a bit of weakness.' (Barrister 5) If a barrister is unlikely to seek help with work from colleagues, it is additionally unlikely that they will want to appear to be unable to cope with other demands of the profession by admitting to having difficulties with wellbeing or mental ill-health.

One of the central issues relating to gender inequality includes how barristers may find it difficult to practise and have a family life, including children. This has previously been observed in the literature

---

70 Collier, 'Wellbeing in the legal profession' (n 6 above).

71 Ibid; Hilary Sommerlad, '"A pit to put women in": professionalism, work intensification, sexualisation and work–life balance in the legal profession in England and Wales' (2016) 23 International Journal of the Legal Profession 61.

72 Jones et al (n 8 above) 74–75.

with respect to solicitors,[73] but the pressures within a solicitor's firm are quite different, as it involves managers placing demands on employees, whereas, for barristers, the demands are created by the need to be able to run a successful practice and/or stem from other colleagues or judges. One participant observed that the successful female barristers she knows often have no children, the implication being that they can dedicate more of their life to work.

Other participants referred to taking maternity leave as being problematic, because they would have to take time out of the profession, and this would affect their relationships with instructing solicitors:

> When you come back after nine months of effectively not working, you have to almost start at the bottom again. There are some people who are lucky enough that they don't have to. They just maintain good enough connections that they don't have to. A lot of people would not do that. A lot of people would be available for work within a matter of weeks because the risk is that you lose the momentum. (Barrister 5)

One participant compared this situation unfavourably with the system in England and Wales:

> Presumably, if you took time off from chambers, your clerks would get you back up to speed with some work as soon as you came back. That doesn't happen here. (Barrister 10)

However, this same issue was recognised in Hunter's work concerning the Australian Bar (where the chamber system is used). Here, Hunter describes how one of the participants raised her gender when discussing returning to work soon after having a child, and explained she tried to hide she was doing this in case she lost work as a result.[74] One participant reflected on the fact that the gender disparities at the Bar also led to similar inequalities within the judiciary (who are often appointed from practising barristers). They indicated that women were more likely to take salaried judicial positions in the lower courts, but overall 'there are fewer female QCs, and fewer female judges' (Barrister 10).

One of the central issues discussed by many of the participants was how barristers are pigeon-holed into specific areas of practice based upon gender. The consensus amongst those interviewed was that female barristers were pigeon-holed into family law, whereas areas like criminal law and commercial law tend to have more men working in them. This was because these areas are perceived to be more 'masculine' in nature. This has been observed in other jurisdictions – in Melville and Stephen's study of Scottish advocates, female advocates were far more likely to be found to specialise in family, child, education and

73 Sommerlad (n 71).

74 Rosemary Hunter, 'Talking up equality: women barristers and the denial of discrimination' (2002) 10 Feminist Legal Studies 113, 120.

personal injury law, where it was perceived that these types of legal work were more suited to supposedly female nurturing skills.[75]

This appears to be perpetuated, at least in part, by how practitioners are instructed. One participant's view was that instructing solicitors see family as the domain of women and other areas of law as the domain of men. Another participant's reflection on this issue was that this is because the same gendered split occurs across the legal profession in Northern Ireland – female solicitors are also pigeon-holed into family law. It might therefore be the case that what is being witnessed is solicitors instructing barristers of their own gender. Another participant commented that it may be due to gendering at a different level – where emotionally impactful family work is perceived as the domain of women, rather than men (the flip-side of this being that men are able to handle criminal cases better because violent individuals are involved and women are supposedly deemed to have nurturing skills). As one participant commented:

> The types of cases that I was instructed in was probably because they wanted a softer female approach. You know, but I wouldn't expect to get landed with, like, a murder brief, for example, because I know that would be going to a male colleague. (Barrister 3)

The same participant highlighted that she felt that male judges could be patronising to female counsel in certain types of cases and that this might affect why a male barrister might be appointed by a solicitor over a female barrister:

> And very much the attitude, of some of the judges, [is] 'look at you like you're a silly wee girl and you know nothing about this, and what are you doing here? ... Could they not have briefed a male barrister?' And, I think there would be a preference of male solicitors that work in them kind of big criminal defence firms to brief male counsel over female counsel. (Barrister 3)

What can be concluded from these discussions is that it is apparent that practice at the Bar is performed, like its counterparts in England & Wales (noted above) and Ireland,[76] in a way which impedes the prospects of women. This suggests there is a need to use the JD-R model to identify, and where possible resolve, specific job demands impacting upon female barristers and also actively seek to build resources to ameliorate such demands, for example, providing targeted training and support for females during and after periods of maternity leave.

---

75 Melville and Stephen (n 55) 218.

76 Kolm Keena, 'Women in law still face a fight for gender equality' *Irish Times* (Dublin, 29 January 2018); Ivana Bacik, Cathryn Costello and Eileen P Drew, *Gender inJustice: Feminising the Legal Professions?* (Trinity College Dublin Law School 2003).

Familial connections

Another inequality discussed by participants, which is also linked to the Bar's reliance on instructing solicitors, is the role of familial connections in enabling some barristers to obtain a pupillage and become more easily established in the earlier days of their career. Some of the participants likened this to a form of nepotism:

> I mean there's the classic thing in law, and especially in the Bar, of people who have family connections, you know, people who are the sons of QCs, sons of barristers, sons of judges, inevitably have a big head start. (Barrister 8)

In addition to having familial connections, having financial support from family members was also perceived as being advantageous. For some members of the Bar who did not have the benefit of financial support from their family, this could make a significant difference in whether they were able to get established. This is because, for early career barristers, there will be times when their practice is not financially sustainable and so financial support from elsewhere – either through family, savings, or additional employment – becomes very important. The importance of familial connections suggests that those from a low socio-economic background, who lack family ties within the legal profession, are, like women within the profession, also likely to be adversely affected. Given that this places additional demands upon specific populations, the application of the JD-R model once again necessitates a consideration of the resources that could be provided to either ameliorate these or provide additional, targeted resources, for example, through enhanced bursary schemes which could help those with low socio-economic standing to access financial support when trying to develop a practice.

THE PRACTICES AND CULTURE OF THE BAR LIBRARY

Institutional practices

Several specific institutional practices relating to the Bar Library were discussed by participants. A number of issues were raised, including the need to 'hot desk' until a permanent desk could be arranged (for which there is a waiting list), or sometimes there being nowhere to sit due to a lack of capacity. The growth of the Bar Library was also referred to, with one participant suggesting it had led to increased competition for work, while another viewed it as increasing diversity. One participant also commented critically on the Bar Library's combined representative and regulatory function, meaning disciplinary proceedings were conducted by peers.

None of the participants considered that the Bar Library system was not an appropriate system for them, with some participants directly referencing the benefits of the collegiate system, or the body of knowledge and experience the Bar Library holds within it. At the same time, some participants viewed potential benefits in aspects of having a chambers system. Participants perceived the main benefit of chambers as being increased administrative support, although one also felt this would be accompanied by increased pressure and scrutiny. Views upon whether a chambers system would increase or decrease levels of collegiality also appeared to vary, meaning overall the arguments for and against reform appeared very finely balanced:

> To be honest, there's something I like about the Bar Library. It does have a bit of a collegial atmosphere. It's perhaps not as tight as it would be in chambers, but you do find a lot of decent people. There are people that you could go to if you needed help with anything. So, I don't think I would really like a chambers system, myself, that much. But then perhaps it would prevent you from taking on too much work because, I think, with the chambers system, you know, the clerk assigns work to people … . (Barrister 8)

Bar Library culture

The issue of collegiality, and how it might be eroded by competition in an adversarial system where the culture is one of practitioners competing, was discussed above. Similarly, the discussion of long and unsociable working hours and the expectation of constant availability also appears to be not only structural, but also cultural. For example, one participant referred to there being an 'unwillingness' to take breaks, whether for half a day or a month, because of the 'stigma' involved in doing so (Barrister 1).

The stigma surrounding mental health and wellbeing issues was also referred to by several participants. Some indicated that the stigma had lessened, or even dissipated entirely, within recent years (citing the Bar Library's annual Wellness Weeks as an example). However, others suggested that it remained, at least to some extent, with two participants suggesting this was a broader societal issue within Northern Ireland. Several participants indicated that emotional wellbeing was not widely discussed at the Bar. They also indicated that there remained a need to project a strong persona or provide a confident performance to remain competitive and retain your reputation:

> People struggle to accept that they may be struggling. And particularly for barristers that are, like, high flying, do you know what I mean, because they think to even admit to yourself that you may need support in terms of emotional wellbeing or mental issues, is nearly like a sign of weakness. (Barrister 1)

This also seemed to be linked to the idea of the Bar as a small community where gossip and rumour could quickly spread, with one participant indicating that they would be reluctant to make even an anonymous complaint against aggressive behaviour from a fellow barrister because of their fear of being identified (Barrister 4). More broadly, this indicates cultural issues which have potentially detrimental impacts upon wellbeing. The remaining stigma around mental health and wellbeing may be a barrier to individuals being able and willing to equip themselves with the emotional and psychological job resources to practise good self-care and reach out for help when required. The indications that such stigma has lessened is encouraging, but it is important that initiatives such as wellbeing weeks run alongside a sustained effort to embed wellbeing at the level of daily practice.

DISCUSSION

The interviews discussed in this paper were conducted prior to the onset of COVID-19. The subsequent global pandemic has had a significant impact upon societal wellbeing internationally, including in Northern Ireland.[77] In terms of the legal profession, although there are potential benefits, including a move to more flexible working patterns, a range of concerns and challenges have also arisen.[78] In July 2020, a report by the Bar Council of England and Wales found that the financial impacts of the crisis, particularly on both publicly funded and early career barristers, were potentially devastating with huge reductions in work and income.[79] It noted that 16 per cent of self-employed barristers wished to leave the Bar, compared to a usual annual turnover of 2 to 4 per cent.[80] Although not expressly discussing wellbeing, it is clear from these findings that the need to address wellbeing issues at the Bar is becoming more, rather than less, important for the long-term sustainability and success of the profession.

Applying the JD-R model to the findings of this study demonstrates that barristers in Northern Ireland face a significant level of job demands, often created (or exacerbated) by the self-employed nature of the role. This brings practical demands, including a potentially heavy administrative burden and the need to undertake multiple roles. These pressures, together with a sense of being in competition with

77 Mental Health and Emotional Wellbeing Surge Cell (COVID-19), 'The mental health impact of the COVID-19 pandemic in Northern Ireland' (Department of Health, 3 August 2020).

78 Richard Collier, 'Wellbeing in the legal profession after COVID 19' (SLSA Blog, 9 September 2020).

79 The Bar Council, 'Bar Survey summary of findings July 2020' (27 July 2020)

80 Ibid note 78.81.

ones' peers for instructions, feed into a workplace environment of long and unsociable working hours, where forms of incivility amongst colleagues are often normalised and work–life balance is too easily disregarded. At the same time, barristers also have access to a range of job resources, including the flexibility of organising their own time, the support of a Bar family and, once established, the potential of work that is both intellectually and potentially financially rewarding.

It appears from this study that job demands are at their highest for early career barristers, namely those still in the process of establishing their career at the Bar. These are exacerbated by a sense that existing legal education and training (including, for some, the experience of pupillage) fails to give a sense of preparedness to allow participants to manage expectations appropriately and gain an understanding of some of the institutional practices at the Bar. It is also apparent that women and other minority populations within the Bar experience additional demands too, particularly in those areas of practice which have traditionally been male dominated. For such individuals, faced with the difficulties of establishing their career in a gendered environment, the impact upon wellbeing is likely to be significant.

To tackle wellbeing issues identified in this study requires a culture where mental health and wellbeing more generally are destigmatised, allowing open discussion and early proactive interventions when required. Where such a culture has not yet developed, there is the danger of losing a more diverse cohort in these early years of practice, something which could have a stultifying effect on the Bar, making it unrepresentative and disconnected from society more widely.

Using the JD-R model, the first question to ask in response to these findings is whether the job demands upon barristers can be ameliorated? This is perhaps particularly problematic where some of the issues, such as the prevalence of self-employment and the role of instructing solicitors, are structural ones, deeply embedded within the functioning of the legal system in Northern Ireland. Although this should not lead to them being treated as sacrosanct, it does suggest that it would be potentially difficult to alter these without significant internal or external pressures to do so. However, alongside these structural issues, there are other job demands where proactive efforts could be made to lessen their effect. This could involve the Bar Library investigating ways of alleviating the administrative burden on members through the provision of some form of clerking service, the introduction of training to assist with working practices, and some form of ongoing dialogue between barristers and with the judiciary around their interpersonal skills.

In addition, as indicated above, it is possible for job resources to have a buffering effect upon the impact of job demands. One of the existing

resources most widely identified by participants was the collegiality of the Bar family model. Considering how to develop this model even further could therefore assist in lessening some of the detrimental impacts of the demands above. This also interlinks with the findings that junior barristers are particularly vulnerable to heightened job demands. A review of the education and training system, in partnership with legal education and training providers and including pupillage, could help identify ways in which to enhance job resources through a particular focus on the wellbeing issues raised (including lack of preparedness). In tandem with this, further consideration should be given to the early years of practice, with a focus on determining (and responding to) the needs of early career barristers, both in practical ways (eg via the provision of more desks or support for home-working) and via emotional and psychological approaches (eg the introduction of formalised wellbeing-support schemes), perhaps emulating some aspects of the Wellbeing at the Bar initiative in England and Wales.[81] Such targeted responses must also be considered in relation to women and other populations whose job demands are exacerbated by current working practices.

An overarching focus of all these initiatives, whether focused on lessening job demands or increasing job resources, must also be on continuing to ensure that the culture of the Bar evolves, so that wellbeing is embedded and destigmatised. This is unlikely to entail a single radical shift, but rather involve a more incremental process of proactive initiatives, formalised opportunities to share good practice and increasingly open dialogue.

CONCLUSION

> It is a profession of survival of the fittest. (Barrister 6)

The perceptions of wellbeing at the Northern Ireland Bar held by participants in this study suggest that it can be a competitive, difficult, even hostile, workplace environment, leading to potentially detrimental impacts on wellbeing. At the same time, some of the forms of collegiality experienced, together with a suggestion of gradually evolving culture, suggest some more positive aspects are present. Applying the JD-R model conceptualises job demands and job resources as a form of balancing act, enabling a holistic appreciation of wellbeing at the Bar. Drawing upon this indicates the potential to ameliorate some job demands and enhance some job resources in a way which is beneficial to wellbeing. To do so will require the engagement and support of key stakeholders to the Bar, such as the Bar Council. The results could

81 The Bar Council, 'Mental health and wellbeing at the Bar'.

move the Bar of Northern Ireland towards a culture where barristers are able to better balance their personal and emotional resources with the demands of the profession – an essential aspiration for a profession which retains such an important status and role within society in Northern Ireland.

In applying the JD-R model, this article emphasises that, where there are demands within the profession (including those affecting specific groups), the Bar Library, the practitioners themselves and other stakeholders can seek to adopt specific mechanisms and push for cultural change to alleviate burdens and enhance wellbeing. This model is useful to conceptualise how practitioners are supported in the Bar in Northern Ireland, and the demands of the profession, but can be used to assess wellbeing in the Bar in other jurisdictions, as well as other legal settings, inside Northern Ireland, and outside of it.

Northern Ireland
Legal Quarterly
Autumn Vol. 72 No. 3 (2021): 480–509
Article DOI: 10.53386/nilq.v72i3.978

No tyranny for failing Donald Trump – sad! Law, constitutionalism and tyranny in the twenty-first century*

Aoife O'Donoghue†

Durham Law School
Correspondence email: aoife.o'donoghue@durham.ac.uk

ABSTRACT

Donald Trump's presidency resulted in several accusations of tyrannical intent. The end of his term of office, and particularly the rioting of 6 January 2021 and the denial of the presidential election results, did little to dispel those accusations. Tyranny, while perhaps not fashionable as a basis of analysis, has a long-intertwined relationship with law and constitutionalism. This article uses Donald Trump's presidency to consider the relationships between tyranny, tyrannicide, law and constitutionalism. The article considers law and constitutionalism's role in both preventing and advancing the advent of tyranny and examines their limitations in stopping tyrannical intent. Public contestation is put forward as an equally significant bulwark against the advent of tyranny, but also a space under tremendous pressure during Donald Trump's presidency.

Keywords: Donald Trump; tyranny; constitutionalism; contestation; impeachment; tyrannicide.

INTRODUCTION

Tyranny is always relevant to debates on political power. In recent years, tyranny has been renewed as an analytical lens in 'the West', even though in reality seeing tyranny as newly relevant is a form of aphasia that ignores long-standing tyrannical regimes in the West (such as Belarus) and foundational tyrannical histories, especially in states with embedded imperial constitutions or founded on racial

† First published in NILQ 72.AD1 (2021) ADVANCE 1 July/August 33–62.

* Professor, Durham University (UK). My thanks to the School of Law, Queen's University Belfast, for the invitation to give the Stephen Livingstone Lecture and for the comments and questions from the audience. My thanks also to the staff of the Law School for the detailed comments at an earlier staff seminar. My thanks to Roger Masterman (Durham University), Fiona de Londras (Birmingham University) and Nicolás Perrone (Universidad Andrés Bello) for their comments upon earlier drafts. Any errors remain my own.

and gender-based exclusions.[1] Nonetheless, the contemporary rise of authoritarianism or illiberal 'constitutionalism' alongside existing long-term tyrannical orders raises concern. Donald Trump embodies these anxieties.[2] This article does not ask whether Trump turned the United States (USA) into a tyranny, but rather whether Trump altered a system, already (always) under pressure, to instigate the processes of becoming a tyranny.

Tyranny cuts across governance structures and easily adapts to its context. No two tyrannies are the same; tyrants adapt tyrannical forms like silence, rule by law, fear, misogyny and illegitimacy to suit the context in which tyrants seek to maintain authoritarian rule. Using a taxonomy of tyranny, the article considers the collation of relational and shifting elements via an aggregation of events, practices and processes which indicates the presence of tyranny. Tyranny, like democracy or oligarchy, is an old concept and like these others, is as potentially present today as it was in Greek city states.[3] Law, tyranny and constitutionalism are intimately linked. Constitutionalism and law hold dual functions, as tools of tyranny and buffers against tyranny's emergence. This mirrors law's duality as both a preventative mechanism and a device to perpetuate harms. Within these interactions, political contestation plays an essential role. Political contestation is vital to both tyranny's prevention and to its removal. The collective political action of a population is more significant to staving off or overthrowing tyranny than constitutional/legal forms. Attempts to blame the 74 million individuals who voted for Trump, even those that stormed the Capital, forget the circumstances, including institutional acquiescence, that led to his candidacy, the forms his presidency took and what has happened to the US public political space.[4]

Trump's clear tyrant-like approach makes tyranny an obvious choice as a lens through which to seek to understand his presidency and its broader significance. However, the analytical power of tyranny is such that it illuminates not only Trump and Trumpian politics as tyrannical, but also enables us to undertake a broader analysis that brings fresh perspectives to contemporary authoritarianism. Tyranny

---

1 S Pahuja, *Decolonising International Law Development, Economic Growth and the Politics of Universality* (Cambridge University Press 2011) 4; N Tzouvala, *Capitalism as Civilisation: A History of International Law* (Cambridge University Press 2020) 2; E W Said, *Culture and Imperialism* (Vintage 1994).

2 T Snyder, *On Tyranny: Twenty Lessons from the Twentieth Century* (Tim Duggan Books 2017) 9–13.

3 S Lewis, *Greek Tyranny* (Phoenix Press 2009) 31.

4 Federal Election Commission, Official 2020 Presidential General Election Results; E Hanley and D Smith, 'The anger games: who voted for Donald Trump in the 2016 election, and why?' (2018) 44 Critical Sociology 195.

as a phenomenon is adaptable; it shifts form to the political system in which it is sought and enacted. US exceptionalism is such that its long history of deploying tyranny beyond its borders is generally recognised, but its internal tyranny is less commonly named as such. However, it is this US exceptionalism, and the USA's place in 'the West', that makes it important for us to turn the analytical lens of tyranny on Trump as well as on other 'obvious' candidates such as Duterte, Orban, Putin and Modi. In all cases, these tyrannical leaders seek to sustain a veil of legitimacy through claims of redefining constitutionalism, attempting to distinguish themselves from 'known tyrants' like Lukashenko or Kim Jong-un. Centring the USA and Donald Trump in an analysis of tyranny reorients a debate that is too often, and inaccurately, orientalised.[5]

This article considers the ability of constitutions to withstand tyrannical aspirations and what other sources of political practice, in particular public contestation, are required to keep tyranny at bay. Like all concepts, tyranny only ever gives a partial account, and this piece does not exclude other explanations of events in the USA or elsewhere.[6] Rather, in focusing on the relationships between tyranny, law and constitutionalism, it considers contemporary constitutional governance and its utility in preventing tyranny. This article begins by discussing tyranny, before examining whether its tendrils are identifiable in the US system.[7] The piece then examines Trump's presidency before considering what lessons may be learned from his time in office.

TYRANNY

Definitions of tyranny are rare, nonetheless a taxonomy of its characteristics is possible and includes illegitimacy, silence, rule by law, fear, gendered and bureaucratic governance which may be beneficent, but the ultimate benefits always come to the tyrant and their cadre. Identifying tyranny requires sufficient latitude so that regimes qualify but with appropriate specificity to not encompass mere poor governance. Tyranny is about more than Nero, Stalin, cruelty or conspicuous consumption. While these are bad, the specific badness is

5 E W Said, *Orientalism: Western Conceptions of the Orient* (Knopf Doubleday 2014) 31–49; A Çirakman, 'From tyranny to despotism: the enlightenment's unenlightened image of the Turks' (2001) 33 International Journal of Middle East Studies 49.

6 See, for instance, C Salmon, 'Trump, fascism, and the construction of "the people": an interview with Judith Butler' 29 September 2016.

7 Bonnie Honig argues there is a rich strain of monarchism and tyranny within the US Constitution: Bonnie Honig, *Shell-Shocked Feminist Criticism after Trump* (Fordham University Press 2021) 5.

not what concerns scholars. On occasion tyranny is dismissed as mere incompetence or malevolence. While it takes such forms, they are not definitional. Illiberal constitutionalism, absolute constitutionalism, authoritarianism, Bonapartism, Caesarism or even totalitarianism are forms of tyranny adapted to their political contexts.[8]

Google Books' Ngram gives a partial history of tyranny's invocation.[9] As a word, 'tyranny' peaked in use in 1785, just before the Federalist Papers were published, followed by a steep decline before rising again from the mid-1980s onwards. 'Tyrant' has two peaks, 1589 and 1787, and then follows a similar pattern.[10] Bonapartism and Caesarism emerge in the early 1800s and remain more popular than tyranny; totalitarianism, while briefly cited in the late 1700s, increased dramatically in the early 1920s before steadily declining after 1949 and then increasing again after 1979.[11] The outlier is authoritarianism, whose use begins in the early 1900s and remains on an upward trajectory.[12] Authoritarianism is evidently the twentieth and twenty-first-century word of choice for overwhelming power.

Descriptive choices vary across political events, but overweening power is a constant source of debate. Specific invocations of tyranny as a distinct concept are valuable not because other words have different meanings, but because of tyranny's specific relationship with law, constitutionalism and legitimacy.[13] Tyranny comes in many modes, informed by the political, cultural, social, and legal context in which it arises, and other terms may be useful in describing, for instance, personality-driven power – Bonapartism or Caesarism – an ideological-driven structure encompassing the entirety of life – totalitarianism – or a form of overbearing power seeping through all power structures within a society – authoritarianism. None of these divisions are strict and, rather, fluctuate across contexts. Tyranny encompasses each of these modes, but also brings debate back to the liberal construction of whether law and constitutionalism may minimise it.

Tyrannical debate, by specifying a groups' proclivity for tyranny, informs how we construct groups, be they women, majorities or

8 M Richter, 'A family of political concepts: tyranny, despotism, Bonapartism, Caesarism, dictatorship, 1750–1917' (2005) 4 European Journal of Political Theory 221.

9 Google Ngram search for 'tyranny'.

10 Google Ngram search for 'tyrant'.

11 Google Ngram search for 'totalitarianism'.

12 Google Ngram search for 'authoritarianism'.

13 Arendt would disagree, H Arendt, *The Origins of Totalitarianism* (Harcourt 1951) 6.

those under Eurocentric gazes.[14] Tyranny underpins how liberal constitutionalism disregards certain groups, like minority elites, as problematic by minimising their proclivities towards power.[15] A long-held idea that all are tyrants in waiting emerges from Western gendered, religious, cultural and historically specific ideas of power and Aristotelian political action where humans tend toward rather than choose roles.[16] But doing so substitutes character or gender types for specific actions. Tyranny has global reach. It evolved at distinct moments across multiple geographic spaces, and, while the focus here is on Western political theory, it is not a western innovation. Second, Western imperialism (including its contemporary form) used and exported both its variety of tyranny and liberal constitutionalism.[17]

To fully understand the role of constitutionalism and contestation, tyrannicide, tyrannophobia and tyrannophilia are also important. Tyrannicide is parasitic, relying on tyranny's illegitimacy for its legitimacy and on tyrannicides to not themselves become tyrants. Tyrannicide encompasses a range of actions up to and including violence, but democracy and processes such as impeachment diminish that space.[18] Hobbes devised tyrannophobia to describe those who feared strong sovereigns or sought substantive democracy and argued that it was their foolish fears that lead to dissent and civil war.[19] Locke, conversely, suggested that tyranny is always extant and vigilance necessary to prevent its emergence.[20] Accusing others of foolish fears or pointing to a system's ostensible ability to withstand tyranny while not recognising systemic gender or racial tyranny is an easy silencing tool and partially responsible for the rise of paradoxes such as authoritarian constitutionalism. Tyrannophilia is an old phenomenon, dating to Plato in Syracuse and his ill-fated attempt to create a

14 T Nyirkos, *The Tyranny of the Majority* (Routledge 2018) 1–3; E W Said, *Orientalism: Western Conceptions of the Orient* (Knopf Doubleday 2014) 31–49; L Bradshaw, 'Tyranny and the womanish soul' in T Koivukoski and D Tabachnick (eds), *Confronting Tyranny: Ancient Lessons for Global Politics* (Rowman & Littlefield 2005) 161.

15 R A Dahl, *A Preface to Democratic Theory* (University of Chicago Press 2006) 9–10.

16 C Epstein, *Birth of the State: The Place of the Body in Crafting Modern Politics* (Oxford University Press 2021) 24.

17 L Colley, *The Gun, the Ship, and the Pen: Warfare, Constitutions, and the Making of the Modern World* (Liveright 2021).

18 F L Ford, *Political Murder: From Tyrannicide to Terrorism* (Harvard University Press 1987) 2; S K Brincat, '"Death to tyrants": the political philosophy of tyrannicide—part I' (2008) 4 Journal of International Political Theory 212.

19 T Hobbes, *Leviathan* (Clarendon Press 1909) 253.

20 J Locke, *The Second Treatise on Government* (Tegg & Co London 1823) 201.

philosopher tyrant, but named by Mark Lilla.[21] A tyrannophile is a theorist who considers a specific tyrannical regime as an avatar to fulfil their own theoretical ambitions, which align with the regime's creed. The theorist allies with the tyranny to see their ambitions fulfilled.[22] Tyrannicide, tyrannophobia and tyrannophilia are tied to each other and form part of a tyrannical political landscape.

TYRANNY, LAW AND CONSTITUTIONALISM

Two elements are critical to preventing modern tyranny: contestation and substantive constitutionalism. Montesquieu, Rousseau, the Federalists and Locke were the first to suggest that modern constitutionalism offered the possibility of preventing tyranny.[23] Arendt showed that the politically active individual is essential, and Dahl argues that polyarchal democracy, a combination of popular sovereignty, political equality and majority rule, is key.[24] But to be politically active one has to access the political sphere, which includes access to information and to knowledge of your governance structures, knowledge of what your constitution says and access to facts/truth. Arendt, Dahl and Machiavelli agree that, while legal formulations are important, what is essential is a network of habits and attitudes held by a politically active individual. In other words, it is the nexus between constitutionalism and the politically active system that prevents tyranny.

When prevention fails, tyrannicide comes to the fore. The legitimacy and duty to remove tyrants is rarely repudiated.[25] Three questions are critical: is there a tyranny; whose duty is it to remove the tyrant; and then what actions may be taken? Modern constitutionalism gives partial answers. The Federalists sought to constitutionalise it through regular elections and formalised processes like impeachment which give constituted power-holders methods to tackle the (potential)

---

21 M Lilla, *The Reckless Mind: Intellectuals in Politics* (re-issue, New York Review of Books 2016) 210–211.

22 T Koivukoski, 'The education of a tyrant' in Koivukoski and Tabachnick (n 14 above) 197; M Kenny 'Reckless minds or democracy's helpers? Intellectuals and politics in the twentieth century' (2004) 3 Contemporary Political Theory 89.

23 Colley (n 17 above); A Rana, *The Two Faces of American Freedom* (Harvard University Press 2010) 3; M Astell, *Political Writings* (Cambridge University Press 1996).

24 R A Dahl, *Democracy and its Critics* (Yale University Press 1989) 27–28, 18; H Arendt, 'The great tradition: I. Law and power – Hannah Arendt's centenary: political and philosophical' (2007) 74 Social Science 713–726, 722.

25 J Canning, *A History of Medieval Political Thought* (Routledge 2005) 113.

tyranny.[26] Other common constitutional features, such as judicial review or term limits, also forestall tyrannical ambition. A threshold for violent rebellion or revolution remains, though it is increasingly of last resort, but one held by constituent power-holders and recognised in legal forms such as remedial or anticolonial self-determination or anti-apartheid movements.[27] Albeit, this does not always legitimate violence by either the constituents (though it does not rule it out) or by the international community entering into humanitarian intervention to 'save' a population.[28] The nexus between the constituent and the tyrant is critical: it is their right and their duty to remove the tyrant, and it is for them to determine the legitimate means of doing so.[29]

However, constitutionalism's relationship with tyranny far predates the Federalists, or the theorists of pre-revolutionary France.[30] This longer history points to the ways in which tyranny emerges from and through law. Tyranny emerged in archaic Greece, and by the classical period its links to legitimacy and law were well embedded.[31] The early emergence of tyrannicide and legal immunities in Greek city states further attest to their close relationship.[32] In Rome, Sulla, Julius Caesar and Augustus sought as leaders to hollow out their political offices and achieved this partially through law.[33] Following Julius Caesar's assassination, Cicero argued that he had held illegitimate authority, an argument followed by the Roman Senate as it outlawed the office of dictator.[34] During the European medieval era, associations between il/legitimacy and tyranny/tyrannicide became bound to

26 US Constitution, art II, s 4; D George, 'Distinguishing classical tyrannicide from modern terrorism' (1988) 50 Review of Politics 390, 390, 407.

27 Neelam Srivastava, 'Towards a critique of colonial violence: Fanon, Gandhi and the restoration of agency' (2010) 46 Journal of Postcolonial Writing 303; A Anghie and B S Chimni, 'Third World approaches to international law and individual responsibility in internal conflicts' (2003) 77 Chinese Journal of International Law 2; B Rajagopal, *International Law from Below: Development, Social Movements and Third World Resistance* (Cambridge University Press 2003); J Dugard and J Reynolds, 'Apartheid, international law, and the occupied Palestinian territory' (2013) 24 European Journal of International Law 867.

28 C R G Murray and A O'Donoghue, 'Towards unilateralism? House of Commons oversight of the use of force' (2016) International and Comparative Law Quarterly 305.

29 A Orford, *Reading Humanitarian Intervention: Human Rights and the Use of Force in International Law* (Cambridge University Press 2003).

30 Colley (n 17 above).

31 M Lane, *Greek and Roman Political Ideas* (Penguin 2014) 75.

32 D Teegarden, *Death to Tyrants! Ancient Greek Democracy and the Struggle against Tyranny* (Princeton University Press 2013) 5.

33 J Rich, 'Making the emergency permanent: auctoritas, potestas and the evolution of the principate of Augustus' (2012) Des réformes augustéennes 37–121, 43.

34 Cicero, *On Duties* (Cambridge University Press 1991) 96.

theocratic monarchs and the question of who possessed the right and duty to overthrow a tyrant, if such a right and duty existed.[35] While Bartolus and Machiavelli framed their debates on who held the right to make law and whether it was possible to be legitimately above the law, the nexus between law and tyranny never lost its centrality.[36]

Across these writings, illegitimacy emerges across three different areas, all of which are concerns of law and constitutionalism, namely, coming to office, altering offices while in power or the use of power outside of constitutional structures – for instance, leveraging economic power or non-state violence. These are not always referred to as tyranny, although they should be. Strauss argues that we are incapable of naming a tyranny, and descriptions of regimes as post-constitutional or illiberal rather than tyrannical attest to his claim.[37] Terms such as authoritarian or illiberal constitutionalism form part of the legal veil tyrants use to cloak their actions in apparent constitutional/legal legitimacy. Kovács and Tóth rightly refer to contemporary European authoritarianism, in which regimes lay claim to being both constitutional and tyrannical, as 'constitutional barbarism'.[38] However, understanding the misnaming of tyranny as a form of constitutionalism requires us to recognise that the fissures which allow such misnomers to continue are often rooted in claims that democratic legitimacy is unnecessary to constitutionalism; a claim that is found both in states with shorter records of substantive democracy and those with imperial constitutions.[39] Such equivocation towards the necessity of democracy to constitutionalism creates space in which authoritarian or illiberal regimes that violate other normative constitutional norms can easily fit and seek to disguise their tyranny in plain sight. Decoupling constitutionalism from democracy enables the so-called 'secret tyrant' who maintains an outward appearance of

35 C Nederman, 'Tyranny' in *Encyclopaedia of Medieval Political Thought* (Springer 2011) 1347.

36 Bartolus of Sassoferrato, *On the Tyrant* (J Robinson (trans) 2012); G Pedullà, 'Machiavelli's Prince and the concept of tyranny' in N Panou and H Schadee (eds), *Evil Lords: Theories and Representations of Tyranny from Antiquity to the Renaissance* (Oxford University Press 2018) 191, 207.

37 L Strauss, *On Tyranny including the Strauss-Kojève Correspondence*, V Gourevitch and M S Roth (eds) (University of Chicago Press, 1961) 22.

38 Kriszta Kovács and Gábor Attila Tóth, 'The age of constitutional barbarism' (*Verfassungsblog*, 7 September 2019); Gábor Attila Tóth, 'Constitutional markers of authoritarianism' (2019) 11 Hague Journal on the Rule of Law 37.

39 Karolina M Milewicz, *Constitutonalizing World Politics: The Logic of Democratic Power and the Unintended Consequences of International Treaty Making* (Cambridge University Press 2020) 26–27; C Murray and T Frost, 'The Chagossians' struggle and the last bastions of imperial constitutionalism' in S Allen and C Monaghan (eds), *Fifty Years of the British Indian Ocean Territory: Legal Perspectives* (Springer 2018) 147.

legitimacy while turning the system to their benefit.[40] At the same time, tyrannophobia prevents direct accusations of tyranny for fear of looking foolish, as Hobbes would have it. Tyrants understand this. They perceive value and importance in a specific form of governance – constitutionalism – steeped in law and only emerge from the cloak of legitimacy afforded by law and constitutionalism when their superficial commitment no longer serves their purposes.

This relationship between law, constitutionalism and tyranny is bolstered by (neo)imperialist international habits of naming some states as 'failed or outlaw states', with attendant loss of status, and others as rearticulations of constitutionalism with continued inter-state esteem and maintained status in and beyond law.[41] International law asks only whether there is a government and not its form. Constitutionalism and statehood are not synonymous, nonetheless a line between states with dispersed governance as automatically negative and tyrannies as acceptable is evident.[42]

Within contemporary tyranny, constitutional structures often remain intact. There may be separation of powers, but the question is whether there is a substantive division in terms that produces a check on constituted power.[43] A simulacrum of division and check is the most common iteration of modern tyranny. There may be judges or legislators, but if their touchstone is the whim of the ruler rather than the demos or constitution, they are a sham check. Contemporary tyranny can also emerge from temporal and functional changes to political office. Office-holders begin to accumulate additional functions, and there are often creative reinterpretations of the offices of president and prime minister to shift power between the two. This is sometimes accompanied by constitutional changes undertaken in the absence of contestation from other constituted power-holders, or within a denuded public sphere. These are neither 'post-constitutional' or newly formed constitutional spaces.[44] To accept the latter in particular is to accept that a tyrant can create legitimacy through their own tyrannical action.

Even though law is everywhere in tyrannies, in these contexts, law's creation, enforcement and adjudication cannot meet even thin conceptions of the rule of law. While the rule of law often has

40 Bartolus of Sassoferrato (n 36 above) 4–5; C de Montesquieu, *The Spirit of Laws*, Thomas Nugent (trans) (Batoche Books 2001) 322.

41 Kovács and Tóth (n 38 above); Tóth (n 38 above).

42 Montevideo Convention on the Rights and Duties of States 1933; T Ginsburg, 'Authoritarian international law?' (2020) 114 American Journal of International Law 221.

43 C Möllers, *The Three Branches: A Comparative Model of Separation of Powers* (Oxford University Press 2013); J N Shklar, *Montesquieu* (Oxford University Press 1987) 85.

44 Strauss (n 37 above) 180.

negative consequences for those not coming within the bounds of liberal constitutionalisms' citizen actors and continues to have a role in (neo)imperialism, nonetheless it retains important functions as regards constituted power-holders.[45] The tyrant is the arbitrator of the common good, combined with an ability to not just be above the law but to be the law's ultimate interpreter of meaning and value. In this setting, law is the will of the ruler and not of the demos and is not subject to contestation. For the tyrant, nurturing public contestation is dangerous, not because, as Hobbes argues, it risks the common good and any demand for it is mere tyrannophobia, rather because for the tyrant it risks making plain that rule by law subsists.[46]

In tyrannies, law creation becomes the tyrant's whim, unfettered from the demos and attuned to the tyrant's own needs. Incrementally, law no longer incumbers political action, creating space for sudden changes to core structures, unconstrained by processes of accountability, alongside slower modifications that over time alter the tenor of the system. As the law serves the tyrant's definition of the common good, the tyrant decides what is necessary, including what constitutes accurate constitutional interpretation and desirable constitutional change. As the ultimate authority and adjudicator of law, the tyrant, in accordance with their decision on the common good, chooses when to follow law.[47] Where this produces positive outcomes, like economic prosperity or the safeguarding of some rights, this emerges from the tyrant's beneficence and the tyrant remains at liberty to change their mind.[48] The common good and the laws needed to establish it remain the tyrant's whim.[49]

Constructing the demos within a tyranny is linked to structure and beneficence, and in particular to majorities and minorities. Though the substantive debate is older, John Adams coined the phrase 'tyranny of the majority', and in the period when democracy first emerges in the

45 Mark Brown, '"An unqualified human good"? On rule of law, globalization, and imperialism' (2018) 43 Law and Social Inquiry 1391; U Mattei and L Nader, *Plunder: When the Rule of Law is Illegal* (Wiley 2018).

46 Hobbes (n 19 above) 253.

47 Ibid 163; H Arendt, *The Origins of Totalitarianism* (Harcourt 1951) 374.

48 H Jones, 'Property, territory, and colonialism: an international legal history of enclosure' (2019) 39 Legal Studies 187; Frank I Michelman, 'Liberal constitutionalism, property rights, and the assault on poverty' (2011) 22 Stellenbosch Law Review 706; David E Bernstein and Ilya Somin, 'The mainstreaming of libertarian constitutionalism' (2014) 77 Law and Contemporary Problems 43.

49 R Boesche, 'Aristotle's science of tyranny' (1993) 14 History of Political Thought 1, 6.

Atlantic World fear of the demos kicks in.[50] Mill, Strauss and others replicate this fear and often combine and (purposively) confound negative populism that merges divisional politics that places blame for economic (or other) hardships on particular groups alongside modern forms of bread and circuses with positive populism that responds to genuine requirements of the masses.[51] In reality, (elite) minorities wielding negative populism is far more commonplace but rarely gets the press that potential majority tyranny receives. Forefronting majorities as innately problematic gives credence to arguments for 10 per cent less democracy.[52] It is essential to not confuse manipulation by elite minorities through negative populism with democratic majoritarianism, for to do so is to blame the masses for the faults of a (tyrannical) elite.[53]

For the tyrant, diminishing individual and collective capacity for contestation, and so the possibility of contesting a tyrant's rise, is essential. Law forms an important element of the production of silence and the diminishing of public contestation. Harmony and tranquillity appear to exist alongside cacophonies of adulation.[54] Violence or its threat produces silence and turns a population's attention toward the private sphere and away from public contestation, making it fearful that the tyrant will use their power, or perhaps their Twitter feed, to set their followers upon the populace. Emptying the public sphere takes time. In Republican Rome, it took a century between Sulla's dictatorship and Augustus's rise to create a new order. Emptying the public sphere takes many forms, from poverty, to fear of economic or status loss, or fear of coming to the attention of the tyrant. Taking decisions away from the public sphere and into the realm of expertise and non-political spheres, to create what Hirsch describes in the legal sphere as a juristocracy, has similar results.[55] Slowly reducing sites of

50 J Adams, *In Defence of the Constitutions of Government of the United States of America* (1787).

51 Nyirkos (n 14 above) 1–3; R Boesche, *Theories of Tyranny: From Plato to Arendt* (Penn State Press 2010) 37–38; C Mouffe, *For a Left Populism* (Verso 2018).

52 Garrett Jones, *10% Less Democracy: Why You should Trust Elites a Little More and the Masses a Little Less* (Stanford University Press 2020).

53 Honig (n 7 above) 142.

54 N Machiavelli, *Discourses on Livy* (Penguin 2013 [1531]) 18–25; H Arendt, *The Promise of Politics* (Schocken 2009) 78; R Boesche, 'Fearing monarchs and merchants: Montesquieu's two theories of despotism' (1990) 43 Political Research Quarterly 741; John of Salisbury, *Policraticus: Of the Frivolities of Courtiers and the Footprints of Philosophers*, C J Nederman (trans) (Cambridge University Press 1990) 231.

55 R Hirschl, *Towards Juristocracy: The Origins and Consequences of the New Constitutionalism* (Harvard University Press 2007); Jones (n 52 above); T Paine, 'Common sense' in M D Conway (ed), *The Writings of Thomas Paine Volume I 1774–1779* (Putman's Sons 1894) 112.

information and the character of that information or access to quality education and producing in its stead an independent 'reality' reduces the capacity for debate and contestation so that, as Arendt argues, 'reality is no longer experienced'.[56] Once reality is corrupted, public contestation becomes impossible.

TRUMP THE TYRANT?

Donald Trump has not yet turned the USA into a tyranny recognisable to liberal constitutionalism. It is not clear that he seeks to be a tyrant, and in truth he, his personality and his wealth are only a partial font of Trumpism. It is vital to not take Trump as anything more than he is: unideological, ridiculous, but successful in achieving the office of the President of the USA. Trump has, however caused much harm, by separating families at US borders, by turning women's stories of abuse into fake news, by undermining Black Lives Matter, and by mishandling the Covid pandemic. These harms build on pre-existing gender, economic and racial harms that the US already grapples with. Trump did not create misogyny and racism but used a misogynist and racist platform of divisional populism for his own ends.[57]

The question, then, must be whether Trump altered a system already under pressure to instigate and/or expedite becoming a tyranny.[58] There are many incidences during Trump's presidency that suggest tyrannical intent and covering them all is unfeasible. Nor is there a 'smoking gun' that catches the Trump administration as wannabe tyrants. Rather, it is a collection of incidences that when brought together expose the possibility of tyranny within an already-denuded political space and with the support or acquiescence of other political figures. Pre-existing problems in the US political-legal structure created a space for Trump's emergence, and it is what Trump did within that space that is important.

Miller argues that:

> Trump's attitude to truth while campaigning was that of a sophist: someone who is indifferent to the truth, using words only to acquire money, fame, and power. When he became president, however, his attitude changed to that of the tyrant … someone who uses power to assert control over the 'truth.' In other words, by the repeated statement of manifest falsehoods, he drew a circle around himself, forcing

56 Montesquieu (n 41 above) 32, 50; J S Mill, *On Liberty and Other Writings*, S Collini (ed) (Cambridge University Press 1989) 75; Arendt (n 47 above) 169.

57 Rana (n 23 above) 3; Salmon (n 6 above).

58 Honig (n 7 above) 6.

> others to choose between submission to his will and recognition of an independent reality.[59]

This is an integral thread of Trump's presidency and, according to the *Washington Post*, the administration made over 33,000 false or misleading claims during his tenure, accelerating in his final months. This alternative reality included whether or not it rained.[60] Institutional duplicity continues both via Trump and by some members of the Republican Party at both state and federal levels.[61] Trump's post-presidency website is a repository of an alternative reality.[62] The extent that the administration could mislead with no repercussions has direct consequences for the public sphere. US law, according to Sunstein, cannot tackle political lying and in particular how political dishonesty operates in the twenty-first century, via social media.[63] The Trump administration used this legal vacuum to create its independent reality to its great advantage.

If reality underpinning discussion erodes beyond typical political positioning, public contestation becomes denuded. There are two realities, including two legal realities. Law is contingent, which makes lies about it easier. Nonetheless, Trumps' relationship to the realities (rather than truth) of US constitutional and wider law played a significant role in what unfolded. Trump made repeated statements about environmental law, trade law, crime, healthcare and abortion access that were inaccurate.[64] These misleading claims created a legal reality where his triumphs and failures were all couched in law that either did not exist or differed from that described. This makes contestation all but impossible: how to debate legal reform if the basis for that discussion is entirely without substantiation. There may be a cacophony of noise, but actual political debate with the Trump administration did not exist.

Trump repeated false claims about the US Constitution, about the Democratic Party's attempt to overthrow the Constitution, the extent of his executive power, the role of the judiciary, impeachment and about the Constitution's history.[65] Trump's 1776 Commission

---

59 P Lee Miller, 'Truth, trump, tyranny: Plato and the Sophists in the era of "alternative facts"' in A J Torres and M B Sable (eds), *Trump and Political Philosophy: Leadership, Statesmanship, and Tyranny* (Springer 2018) 17.

60 Washington Post Fact Check #14, D Trump (21 January 2017).

61 G Kessler, 'Trump made 30,573 false or misleading claims as president. Nearly half came in his final year' (*Washington Post,* 23 January 2021).

62 See 'The 45th President of the United States' The Office of Donald J Trump.

63 C R Sunstein, *Liars: Falsehoods and Free Speech in an Age of Deception* (Oxford University Press 2021).

64 The *Washington Post* has a searchable database: Fact Checker.

65 Ibid.

is a rewriting of constitutional, legal, political and social history.[66] The report ignores manifest injustice, or slavery's part in the USA's founding.[67] Rather, the 1776 Commission Report follows Trump's will to create an alternate history where the Constitution weakened rather than entrenched racism. This alternative constitutional history resituates Martin Luther King's protests and the necessity of the Civil Rights movement as fulfilling the Constitution's purposes.[68] The history of constitutional law becomes the history of Trump's will. Constitutional scholarship – legal and historical – should not buttress a system's legitimacy by ignoring how it evolved, including tyranny through slavery or disenfranchisement. By repeatedly attempting to assert a truth about law and the Constitution, Trump forces a choice between his will and independent reality but also delegitimises a history of protest and campaigning critical to US constitutional history. Alternative realities force choices between Trump's legal and political world and independent reality and are essential to understanding other aspects of the administration. In the context of the Black Lives Matter movement, the deracialisation of historic protest directly impacted on perceptions of that movement. They were no longer the heirs to Martin Luther King, but rather are in opposition to his legacy. Such strategies delegitimise the public space for contestation by creating the space for tyranny to emerge.

If tyrants enter office to gain benefits, be that economical, psychological, or ideological, the question becomes: what benefited Trump and his cadre? Two potential benefits come to the fore: economic advantages and immunity. Kuhner describes Trump's crony capitalism during his time in office.[69] Trump's Finance Bills increased the wealth of people who resembled him, the wealth-dominating classes.[70] There were some benefits for the wider population, for instance, employment rose, but Trump also made exaggerated claims about his management of the US economy.[71] Most gains went to those like Trump, already wealthy individuals, whose prosperity increased. There were also many conflicts of interest between his and his family's businesses and the office of president. Citizens for Ethics, a non-partisan corruption

66 The 1776 Report was almost immediately withdrawn from the White House website when the Biden Administration took over.

67 Rana (n 23 above) .

68 R Houghton and A O'Donoghue, 'Manifestos and counter-manifestos: an explainer for the 1776 Commission' (*Critical Legal Thinking*, 21 January 2021).

69 T K Kuhner, *Tyranny of Greed: Trump, Corruption, and the Revolution to Come* (Stanford University Press 2020); for a broader view of tyranny, law and the global economy, see L Wenar, *Blood Oil: Tyrants, Violence, and the Rules that Run the World* (Oxford University Press 2018).

70 Consolidated Appropriations Act, 2018 Pub L 115–141.

71 *Washington Post* Fact Check #30,562; #30,557.

watchdog counted 3400 visits by foreign government officials to Trump-owned properties, including Mar-a-Lago, taxpayer spending at Trump businesses, and Trump's promotion of his businesses.[72] There are also his ongoing battles regarding his tax returns and broader financial dealings, which includes cases involving the House Ways and Means Committee and Trump's refusal to pass documents to it.[73] The US Supreme Court held in *Trump v Mazars* that while separation of powers issues were raised, it was not a question of executive privilege, as Trump claimed, and Congress may request, with legitimate reason, documents from a president.[74]

The second benefit derives from the immunities and pardoning powers of which Trump took full advantage. The presidential power to pardon rests in article II, section 2 of the US Constitution.[75] The suggestion that he would pre-emptively pardon himself and his family before leaving office did not come to fruition, but Trump pardoned several individuals either connected to him and/or who were charged or convicted of financial crimes.[76] This includes Steve Bannon (who is subject to fraud charges), Kwame Kilpatrick (found guilty of racketeering and extortion), Paul Manfort (guilty of financial crimes), Anthony Levandoswki (guilty of stealing trade secrets) and Duncan Hunter (convicted of stealing campaign funds).[77] Others pardoned include four Blackwater guards found guilty of murdering Iraqi citizens. Their pardons potentially violate the USA's international legal obligations.[78] Some pardons were connected to the 2016 presidential election or the administration's early period, including that of Michael Flynn, retired army lieutenant general and former National Security Advisor, guilty of making false statements to US federal investigators, alongside political consultants Roger Stone and George Papadopoulos, both guilty of several crimes regarding Russian involvement in the 2016

72 Citizens for Ethics; Honig (n 7 above) 30–35.

73 Committee on Ways and Means v US Department of Treasury.

74 *Donald J Trump, et al v Mazars USA, LLP and Committee on Oversight and Reform of the US House of Representatives* 591 US 140 S Ct 2019.

75 US Constitution, art II, § 2; B C Kalt, 'Pardon me?: the constitutional case against presidential self-pardons' (1996) 106 Yale Law Journal 779.

76 F O Bowman III, 'Presidential pardons and the problem of impunity' (2021) 23 NYU Journal of Legislation and Public Policy; Lauren Mordacq, 'Fake news: the president has the "absolute right" to pardon himself' (2020) 13 Albany Government Law Review Online.

77 Pardons Granted by President Donald Trump; Ken Kurson, a friend of the Trump family, found guilty of cyberstalking was pardoned.

78 'Issuing several pardons, President Trump intervenes in proceedings of US troops charged or convicted of acts amounting to war crimes' (2020) 114 American Journal of International Law 307; United Nations High Commissioner for Human Rights Press Release, Press Brief on the United States (19 November 2019).

US presidential election. Presidential pardoning had already grown to encompass direct political pardons, and Trump took advantage of this trend to extend it substantially.[79]

Coupled with the near 200 pardons during Trump's final two months in office was the rush to execute prisoners. Excessive punishment and pardoning are often tyrannical traits.[80] At the direction of the executive, 10 people were executed, a break in the tradition of pausing executions before a new President takes office as well as a moratorium on federal executions in place since 2003. The US Supreme Court denied three stay requests and a stay of execution granted by the 7th Circuit Court of Appeals.[81] Going ahead with such executions is a call back to practices of bread and circuses in an all too gruesome fashion; executions to appease a base of voters and an inhumane benefit of the common good. It also reveals the weakness in constitutional conventions, especially those related to protecting human rights, here the fundamental right to life, in the context of executive power. Siegal argues that the 'disregard of political norms that had previously constrained presidential candidates and Presidents, and his flouting of nonlegal but obligatory "constitutional conventions" that had previously guided and disciplined occupants of the White House' was one of the most troubling parts of his administration.[82] It shows the vulnerability of relying on such political/legal norms to reign in an administration that can simply choose to ignore them.

Connected to the pardoning/execution axis is the role of the Department of Justice and the Executive Branch regarding investigations. Early in the Trump administration there was a high proportion of investigations into Trump's associates, which slowly lessened, particularly after William Barr became Attorney General. James Comey's firing as head of the Federal Bureau of Investigation (FBI), Bowman describes as an 'emasculation of an executive and congressional mechanism for investigating, and where appropriate punishing misconduct by the president and his allies'.[83] Congress, civic watchdogs and state justice systems continue to investigate, though the Trump administration refused to cooperate.[84] For example, there are two congressional investigations into an Inland Revenue Service whistle-blower's allegations that a Trump political appointee at the

79 Bowman (n 76 above).

80 John of Salisbury (n 54 above) 51–55.

81 *Barr, Attorney General, et al v Hall*, Orlando Order List: 592 US.

82 Neil S Siegel, 'Political norms, constitutional conventions, and President Donald Trump' (2018) 93 Indiana Law Journal 177.

83 Bowman (n 76 above).

84 *Donald J Trump v Cyrus R Vance, Jr* 140 S Ct 2412; 207, District Attorney of New York Statement.

Treasury Department interfered with the audit of President Trump or Vice President Pence.[85]

Several features can help us to understand the tyrannical context at play here. First is the relationship with law and constitutional enforcement as between the three branches of US Government, a point returned to regarding impeachment, the judiciary and the production of silence. Former FBI director Comey's firing made headlines, but there were other many replacements amongst the agencies of the executive branch, particularly amongst those with investigative powers.[86] The very public and humiliating firing of Comey, among others, produces fear of loss of position, a privatising of the administrative sphere of government which increasingly contained people holding back for fear of losing their jobs, or those happy to acquiesce to the Trump administration's desires or on board with Trump's objectives. This was accompanied by what Honig describes as Comey's feminisation, including during Senate Intelligence Committee hearings which used gender as an 'apparatus of power' to undermine Comey's credibility.[87] The apparent reduction in Department of Justice investigations also suggests the possibility that there was an understanding that the Trump administration's need to get its agenda done – the Trump view of the common good – was more important than investigating potential crimes or malpractice in office.

Removing people who voiced opposition or obstructed their agenda was important to the administration. Comey's firing clarified that, if one disagreed with Trump, one should fear for one's job. Comey argues that his firing shows he was doing his job properly, while this may not be the measure, it is undoubtedly how both sides regarded the situation.[88] Political appointees are tied to their appointer, but these offices include independent functions inconvenient to Trump. Yet, there is little within the constitutional or legal structure which prevented the silencing and privatising of these spaces, partly because the very intention was to prevent the 'wrong' investigations from taking place. Nonetheless, political contestation withstood some incidences,

85 US Congress, House Ways and Means Committee, *Donald J Trump, et al v Mazars USA, LLP and Committee on Oversight and Reform of the US House of Representatives* 140 S Ct 2019.

86 M Quinn, 'The internal watchdogs Trump has fired or replaced' (*CBS NEWS*, 19 May 2020) (listing five inspectors general – Intelligence Community, Transportation Department, Defense Department, Department of Health and Human Services, and State Department – fired or replaced by President Trump as of 19 May 2020); B McCarthy, 'Trump has pushed out 5 inspectors general since April. Here's who they are' (*POLITIFACT*, 19 May 2020).

87 Honig (n 7 above) 36–22.

88 J Comey, *Saving Justice: Truth, Transparency, and Trust* (Pan Macmillan 2021).

particularly with Anthony Fauci. That Fauci remained in office, despite the Trump administration's continuing downgrading of the National Institute of Allergy and Infectious Diseases advice on Covid, and, of Fauci himself, suggests there were still restraints – though not legal ones – on Trump's actions. Trump retweeted posts with #FireFauci, accused him of multiple mistakes and now indicates he might fire him if re-elected.[89] Fauci's reputation and public confidence made him difficult to fire, proving contestation may succeed where law provides trifling safeguards.[90] If the intention was to create an executive administration in thrall to Trump, it only partially succeeded.[91] The transition team eventually worked with Biden, and the Justice Department did not find any voter fraud to assist Trump's claims of a stolen election.[92]

On first examination, Trump's presidency seems very loud and far from silent. Yet, silencing occurred all the time. US media regulation is often light touch and freedom of expression sacrosanct, but newspapers, television and other news sources are still subject to regulation, with the critical exception of social media.[93] Zick catalogues an 'extraordinary number of incidences in which the statements, actions, and reactions of Donald Trump, first as a candidate and then as president … questioned or threatened First Amendment values and rights'.[94] Several lawsuits regarding Trump's blocking of Twitter accounts and revocations of press credentials for White House correspondents as potential violations of first amendment rights are ongoing.[95]

---

89 Katie Shepherd, John Wagner and Felicia Sonmez, 'White House denies Trump is considering firing Fauci despite his retweet of a hashtag calling for his ouster' (*Washington Post,* 13 April 2020); Josh Lederman and Kelly O'Donnell, 'White House seeks to discredit Fauci as coronavirus surges' (*NBC News,* 13 July 2020); William Cummings and Courtney Subramanian, '"I appreciate the advice": Trump tells crowd chanting, "Fire Fauci!" to wait until after election' (*USA Today*, 2 November 2020).

90 D P Christenson and D L Kriner, 'Does public opinion constrain presidential unilateralism?' (2019) 113 American Political Science Review 1071.

91 Cass R Sunstein and Adrian Vermeule, 'Presidential review: the president's statutory authority over independent agencies' (2021) 109 Georgetown Law Journal 637.

92 M Balsamo, 'Barr says Justice Department found no evidence of fraud that would change election outcome' (*PBS News Hour*, 1 December 2020).

93 Sunstein (n 63 above) 103; S Fish, *There is No Such Thing as Free Speech* (Oxford University Press 1994); A Meiklejohn, *Free Speech and its Relation to Self-Government* (Harper 1948).

94 T Zick, *The First Amendment in the Trump Era* (Oxford University Press 2019) xii.

95 *Knight First Amendment Institute v Trump* Docket nos 1:17-cv-05205; *CNN v Trump* 1:18-cv-02610-TJK.

Social media plays an important role in contestation and especially in a publicly active sphere connecting individuals across networks, but it remains a largely unregulated space, especially with regards to hate speech and false information.[96] Trump adroitly used law's absence, particularly on Twitter, to silence.[97] He used the official Office of the President Twitter account to retweet his personal account, fusing the two entities. Fear of coming to Trump's attention and the Twitter onslaught that might follow had a chilling effect. During the pandemic, Fauci and his family were harassed and received many death threats.[98] Reich describes how Trump used both the presidency and Twitter to silence private citizens, companies, reporters, union representatives, officials like Fauci and, in some circumstances, other branches of government.[99] Reich details how, when an 18-year-old woman who attended a political forum told Trump he was not 'a friend to women', Trump tweeted that she was an 'arrogant young woman'.[100] Immediately, she began receiving violent, threatening messages, including of rape, attempting to silence her or force her to recant. Trump did not tweet such specific threats, but rather used this non-governmental power based within an unregulated space to point his followers to this woman. Trump also pointed Twitter towards members of the judiciary. Judge James Robart gave a rescinding order against the Muslim immigration ban, Trump on Twitter branded him a 'so-called judge' – presumably, a real judge is one that does not contradict the executive – and what followed were similar threats to Robart's safety, many coming from Russia.[101] Trump endorses the threatening and violent acts of his supporters and Russian bots, using the 'mob' which he partially created through divisive populism to achieve silence. This divisional populism turns contestation upon itself and curtails the

96 C R Sunstein, *Republic: Divided Democracy in the Age of Social Media* (Princeton University Press 2018).

97 Alexandra A Siegel et al, 'Trumping hate on Twitter? Online hate speech in the 2016 US election campaign and its aftermath' (2021) 16 Quarterly Journal of Political Science 71; Douglas B McKechnie, '@ POTUS: rethinking presidential immunity in the time of Twitter' (2017) 72 University of Miami Law Review 1.

98 'Fauci admits administration has restricted his media appearances, says he's not surprised Trump got COVID' (*60 Minutes*, 19 October 2020).

99 Robert Reich, 'Trump's creeping tyranny' (8 December 2016). Honig describes a similar outcome when members of the administration would follow Trump's lead to bring down threats and protests for actions that were civil and ordinary: Honig (n 7 above) 59–67.

100 'Donald Trump attacked her on Twitter' (*CNN News*, 9 December 2016).

101 N Colarossi I, 'Russians sent thousands of threats against US Judge James Robart after Trump tweeted about him' (*Newsweek*, 19 February 2021); see also Judge Gonzalo Curiel, D Kellner, 'Donald Trump as authoritarian populist: a Frommian analysis' in J Morelock (ed), *Critical Theory and Authoritarian Populism* (University of Westminster Press 2018) 71, 76.

politically active space by focusing it on hate and blame rather than at those who truly benefit from the system.

Unregulated social media is in stark contrast to the censoring of traditional media by curtailing its access, suggesting that violence against its members is acceptable, or openly mocking reporters with disabilities.[102] Creating an alternative reality is partially achieved through the fake/failing news discourse but also goes back to Trump's trumpeting of the 'birther' conspiracies that undermined Obama's legitimacy as President.[103] The role of social media in spreading such conspiracies is acute. There are now conspiracies that claim Biden is not really the President, that he is filmed on a 'White House' film set and that mass arrests followed his inauguration. These are set alongside QAnon's alien conspiracies which Trump refuses to deny. This delegitimates both the Presidents before and after Trump. Social media and the media generally in the US are increasingly partisan and alternative facts or misleading claims take the place of a contested political sphere; thus silence is created. Social media is not so different from other news sources and can be positive, for instance, sharing information of police brutality, but the absence of regulation accelerates the denuding of the public sphere.

Societal, political and official violence are features of US politics and life. The everyday violence faced by African Americans at the hands of the state is well documented.[104] Trump's close relationship with violence is not exceptional, but the objective of deployment of violence and the source of the violence sets it apart. Some of this emerged from his social media and his public statements. From the outset there was violence at Trump rallies. Protestors were harassed and sometimes assaulted. There are his rally speeches where he stated 'in the good old days this doesn't happen because they used to treat them very, very rough. And when they protested once, you know, they would not do it again so easily' or 'I'd like to punch him in the face I'll tell you.'[105] The violence is also indirect. The ramifications of his insistence on calling the Covid pandemic the 'Chinese virus' are still being felt in the mass increase in hate crime against the Asian American community in the USA.[106]

---

102 Ibid 73–74.

103 'Fact check: White House was not vacant for ten days after inauguration; night-time darkness pre-dates Biden' (*Reuters*, 3 February 2021).

104 D W Carbado and P Rock, 'What exposes African Americans to police violence' (2016) 51 Harvard Civil Rights–Civil Liberties Law Review 159; N A Cazenave, *Killing African Americans: Police and Vigilante Violence as a Racial Control Mechanism* (Taylor & Francis 2018).

105 G Lopez 'Don't believe Donald Trump has incited violence at rallies?' (*Vox*, 12 March 2016).

106 K Yam, 'Anti-Asian hate crimes increased by nearly 150% in 2020' (*NBC News*, 9 March 2021).

Trump's speech on 6 January 2021 and the extremist violence aimed at Congress – another branch of government – resulted in his second impeachment. Trump's political actions on 6 January used non-state violence as its tool. Although those that stormed the Capital were a small subset of his supporters, Trump felt they were a group he could wield against another part of the governance order.[107] Trump's treatment of Vice President Pence in the midst of the speech is illuminating. He mentioned Pence 13 times, and the mob chanted 'hang Mike Pence' as it entered the Capitol, while in the midst of the violence Trump tweeted that Pence did not have the courage to do what he should have.[108] This was a reference to Pence deciding he was legally bound to certify Biden's election as President. Pence chose the law rather than Trump's reality by voting to certify Biden's election, but, arguably, 147 Republicans who voted against it did not. Pence certified and attended the inauguration; therefore, the threat of violence did not work, but it is telling that Trump thought it might.

There is also less directly attributable violence. The far-right rally and counterprotests in Charlottesville are a prime example, but there were other incidences involving the Proud Boys (now declared a terrorist entity by Canada) and other extremist groups.[109] Trump's non-condemnation, insistence that there are 'fine people on all sides' and suggestions that the Proud Boys should 'stand down and stand by', all point to a long-arm use of violence to achieve a political end.[110] When examined in comparison to the state violence deployed against Black Lives Matter protestors, there is a clear distinction between those who are welcome to voice contestation and those who are not. Over the course of the administration there was a process by which violent/military crackdowns became acceptable against those outside Trump's 'fine people' category. Those engaged in contestation and protest were teargassed, while those engaged in extremism were praised. Examining this alongside the use of racist rhetoric against Mexicans, Muslims and Chinese people, all of whom suffer increased hate crime,

107 August H Nimtz, 'The Trump moment: why it happened, why we "dodged the bullet", and "what is to be done?"'.

108 Sidney Blumenthal, 'The martyrdom of Mike Pence' (*The Guardian*, 7 February 2021).

109 Government of Canada, Public Works and Government Services Canada (3 February 2021), Canada Gazette, part 2, vol 155, no 2 'Regulations Amending the Regulations Establishing a List of Entities'.

110 US Presidential Debate (*Corporation for Public Broadcasting*, 29 September 2020); 'Full text: Trump's comments on white supremacists, "alt-left" in Charlottesville' (*Politico*, 15 August 2017).

the divisional populism of the rhetoric is evident.[111] This is intermixed with fear of Muslims and certain majority Muslim countries (but not Saudi Arabia) and an eliding of Islam with ISIS and Al-Qaeda.[112] Fear of violence from extremist terrorism is genuine, however, it was right-wing terrorism and violence that was manifested at the end of Trump's administration.

The violence of family separations at the USA's borders remains an ongoing issue.[113] A zero-tolerance policy that violated US law and intended to dissuade individuals from trying to reach the USA placed children in cages and separated them from their adult carers.[114] This was followed by a failure to reunite families even after court decisions requiring the state to act.[115] There is immediacy and intimacy to violent family separation, particularly when used as a deterrent against people already in desperate situations. Violence against families and children is illegal, but official Trump administration policy.[116]

Trumps' misogyny is pervasive.[117] His behaviour is imbued with toxic masculinity, which included his unwillingness to wear a mask during the Covid pandemic.[118] But he also deploys what Honig refers to as 'ambi-gendering' as a source of political and rhetorical power.[119]Much of the toxicity is performed in his bombast, his militaristic language, in the normalising of locker-room talk or his discussion of journalist Meghan Kelly '[y]ou could see there was blood

---

111 D A Graham, A Green, C Murphy and P Richards, 'An oral history of Trump's bigotry' (*The Atlantic*, June 2019).

112 *Washington Post* Fact Check #29,317; A Ayoub and K Beydoun, 'Executive disorder: the Muslim ban, emergency advocacy, and the fires next time' (2016) 22 Michigan Journal of Race and Law 215.

113 J Todres and D Villamizar Fink, 'The trauma of Trump's family separation and child detention actions: a children's rights perspective' (2020) 95 Washington Law Review 377; Caitlin Dickerson, 'Hundreds of immigrant children have been taken from parents at US border' (*New York Times*, 20 April 2018).

114 Office of The Attorney General, Memorandum for Federal Prosecutors along the Southwest Border (2018).

115 P Repard, 'Judge says government has 'sole' duty to find, reunite immigrant parents, children' (*San Diego Union-Tribune*, 3 August 2018); see also Order Granting Classwide Preliminary Injunction, Ms L, 310 F Supp 3d at 1145.

116 Todres and Fink (n 113 above) 384

117 Kellner (n 101 above) 71, 76

118 'Trump tells allies his wearing a mask would "send the wrong message," make him look ridiculous' (*NBC News*, 7 May 2020); C Palmer and R Peterson, 'Toxic mask-ulinity: the link between masculine toughness and affective reactions to mask wearing in the COVID-19 era' (2020) 16 Politics and Gender 1044.

119 Honig (n 7 above) 42, 139; P Elliott Johnson, 'The art of masculine victimhood: Donald Trump's demagoguery' (2017) 40 Women's Studies in Communication 229; Maria Aristodemou, 'Unravelling the Trump virus: on Honig's shell shocked' (*Critical Legal Thinking*, 30 March 2021).

coming out of her eyes, blood coming out of her wherever'.[120] John Bolton, his former National Security Advisor, suggests that Trump had trouble with women leaders such as Theresa May and Angela Merkel.[121] While this does not necessarily set him apart from others, his willingness to be publicly and explicitly misogynistic as President (and as a candidate) emboldens others to give voice to their misogyny while negatively impacting upon women's ability to exist in the public sphere and their capacity to call out others for their behaviour.[122] His behaviour did galvanise women's marches and social media campaigns, but it pushed women to retake ground that many assumed won. It also demonstrates the fragility of women's voices and place in the political sphere and their believability, as demonstrated by the 25 women who directly accused Trump of sexual assault or Christine Blasey Ford's testimony against Supreme Court appointee Brett Kavanaugh.[123] During the #MeToo movement, where women's believability was being trumpeted, the Trump administration pushed the opposite agenda.[124]

Evangelical politics has been a feature of the US system since at least the 1970s, and, gradually, politicians have aligned their legal political choices with evangelical policy.[125] For women's rights and particularly their bodily autonomy in the US and globally, this is of immense concern, especially as it impacts on US aid programmes.[126] Trump's appointments to the judicial benches may have their long-term impact here should US federal courts increasingly align themselves with changes to law which curtail women's bodily autonomy or increase the possibilities of discrimination against particular groups.[127] There

120 D P McAdams, *The Strange Case of Donald J Trump: A Psychological Reckoning* (Oxford University Press 2020) 120; J H Rhodes et al, 'Just locker room talk? Explicit sexism and the impact of the access Hollywood tape on electoral support for Donald Trump in 2016' (2020) 37 Political Communication 741.

121 'Trump "has trouble with women leaders", says ex-adviser John Bolton' (*Sky News*, 25 May 2020).

122 Salmon (n 6 above).

123 M K Maas et al '"I was grabbed by my pussy and its# NotOkay": a Twitter backlash against Donald Trump's degrading commentary' (2018) 24 Violence Against Women 1739.

124 Honig (n 7 above) 89, 123.

125 A R Schäfer, *Countercultural Conservatives: American Evangelicalism from the Postwar Revival to the New Christian Right* (University of Wisconsin Press 2011); Charlie Jeffries, 'Adolescent women and antiabortion politics in the Reagan administration' (2018) 52 Journal of American Studies 193; C Gustavo Poggio Teixeira and J Felipe Ribeiro Calandrelli, 'Donald Trump and neoconservantism' (2017) 24 Esboços 380.

126 T McGovern, 'From bad to worse: global governance of abortion and the Global Gag Rule' (2020) 28 Sexual and Reproductive Health Matters 54.

127 A R A Aiken, 'Erosion of women's reproductive rights in the United States' (2019) British Medical Journal 366.

are also concerns as regards LGBTQ rights. The increase in anti-trans rights rhetoric, including banning trans individuals from serving in the military, plays a significant role in gendering who ought to be the active political subject of US politics, and who gets to contest within the political sphere.[128]

Whether law became the whim of the administration is of central importance. Trump endeavoured to bend law and policy to his whim even when it was illegal. He attempted to reintroduce waterboarding or worse, and he banned Muslims from coming to the USA.[129] Over the twentieth century there has been increasing use of executive orders, Trump deploying them is arguably a continuation of the trend of extensive executive power.[130] While it may have been an attempt to change the office of President, it reflects a trend, and not necessarily a positive one. The pressure placed on Pence to not certify the election results, to reinterpret the law to make it Trump's will, is a further example of an attempt to reinterpret law, to create another reality in Trump's favour. There are also the attempts to deny the election result and the tactical, though failed, use of law to stop the count, or deny the results. All but a few lawsuits failed. Trump and his supporters focused not on the absence of credible evidence or persuasive legal argument, but rather on a supposed corrupt system. In reality, it is merely a system that had yet to fall under his thrall.

TYRANNOPHILIA AND JOHN BOLTON

Since exiting the Trump administration, John Bolton is vocal in his views of the potential damage that Trump did to US foreign policy and the US Government, but from Bolton's very specific point of view.[131] From his book it appears Bolton also had concerns while part of the Trump administration.[132] John Bolton is a well-established critic of the international legal and political order including of the United Nations,

128 A Feuer, 'Justice Department says rights law doesn't protect gays' (2018) Supreme Court Preview 468; Presidential Memorandum for the Secretary of Defense and the Secretary of Homeland Security, Military Service by Transgender Individuals (25 August 2017) (note that this document was not included in the Federal Register); M J Lang, 'Examining the Trump administration's transgender service ban through an international human rights law framework' (2017) 25 Duke Journal of Gender Law and Policy 249.

129 Honig (n 7 above) 74.

130 D M Driesen 'President Trump's executive orders and the rule of law' (2018) 87 University of Missouri-Kansas City Law Review 489.

131 J Bolton, *The Room Where It Happened: A White House Memoir* (Simon & Schuster 2020).

132 Ibid.

the International Criminal Court and the Iran Nuclear Deal.[133] His time in the Bush Administration aligned with the acceleration of neoliberal policies of military intervention.[134] Ideologically, what Trump represents other than unrestrained capitalism and his own aggrandisement is unclear. John Bolton, nonetheless, saw Trump as an avatar for his own ideological outlook and their views on Iran coincided, though Trump did quip that he had held Bolton back from interventionist actions. It could be asked whether Bolton acted from a form of tyrannophilia.[135] That he regarded the Trump administration as an avenue for getting his own political and philosophical agenda into practice and that he believed he could tame the reckless nature of the administration. But, as all who find themselves in that position, he learned that it is impossible to turn the subject into one's ideal 'philosopher' or, in Trump's case, ideal interventionist tyrant.

TYRANNICIDE AND THE FAILURES OF IMPEACHMENT

Impeachment is partially at least a constitutionalised form of tyrannicide. If there is an aspirant tyrant, impeachment provides a system of halting tyrannical transformation. Impeachment is an invention of English constitutionalism and partly directed at the potential of arbitrary and tyrannical government, albeit it has fallen into abeyance.[136] Impeachment is not always about tyranny. It can be about forms of treason, bribery, high crimes or misdemeanours that have little to do with tyranny.[137] In Trump's case, the second impeachment is closer to an act to remove or prevent the return of a tyrant than the first. Nonetheless, both reveal the possibilities and limitations within the US constitutional and political processes for removing a tyrant.

Both impeachment processes unfolded within the dual realities created by the administration. For instance, there was an insistence

133 J R Bolton, 'Should we take global governance seriously' (2000) 1 Chicago. Journal of International Law 205, J Bolton, *Surrender is not an Option: Defending America at the United Nations* (Simon & Schuster 2008).

134 J R Bolton, 'The risks and weaknesses of the International Criminal Court from America's perspective' (2001) 64 Law and Contemporary Problems 167; Teixeira and Calandrelli (n 125 above).

135 Lilla (n 21 above) 210–211.

136 Erskine May, *Treatise on the Law, Privileges, Proceedings and Usage of Parliament* (Charles Knight & Co 1844); Jack Simson Caird, 'Impeachment' (House of Commons Library CBP7612 2016); F O Bowman III, *High Crimes and Misdemeanors* (Cambridge University Press 2019) 22.

137 For classical variations on similar processes of account, see M Ostwald, 'The Athenian legislation against tyranny and subversion' (1955) 86 Transactions and Proceedings of the American Philological Association 86.

that the first impeachment was a hoax and illegitimate.[138] Trump argued he no longer recognised House Speaker Nancy Pelosi and that Congressman Adam Schiff, one of the main investigators, should be arrested for treason.[139] While it is natural that those subject to impeachment would argue that it is wrongfooted, the extension of this to undermining the entire system and the corresponding acquiescence of Republicans within Congress established a narrative that passed into the second impeachment which Republican Senator Josh Hawley called a Kangaroo Court.[140] While neither impeachment was successful, the continued undermining of integrity of the process weakens any future attempts to use it to remove a president from office, and thus of a key potential structural guard against tyranny.

The second impeachment shows the cumulative effects of everything that had gone before, from the first impeachment to the silencing, to fear, to the dual (legal) realities, as well as the threats of violence should have Trump lost the 2016 election.[141] Incitement of insurrection and Trump remaining 'a threat to national security, democracy, and the Constitution if allowed to remain in office, and [that he] has acted in a manner grossly incompatible with self-governance and the rule of law' is an exceptionally disturbing indictment.[142] The articles of impeachment referred to both the events of 6 January and a telephone call that 'urged the secretary of state of Georgia, Brad Raffensperger, to "find" enough votes to overturn the Georgia Presidential election results and threatened Secretary Raffensperger if he failed to do so'.[143] This latter element was part of a year-long campaign, begun well before the election itself, asserting that the election results were inevitably fraudulent if Trump lost. The article of impeachment touches on several elements of the taxonomy of tyranny: the use of fear by way of violence and threats to Secretary Raffensperger; the creation of silence via misleading claims and forcing individuals and groups to accept the Trumpian reality as fact; the undermining of elections to overstay a term in office; and an attempt to create an order the ultimate beneficiary of which is Trump. Some House Republicans accepted that

138 *Washington Post* Fact Check #23,564, D Trump, 'When the fake impeachment happened, it was a total fake' (17 September 2020); Fact Check #15,865, D Trump (2 January 2020).

139 *Washington Post* Fact Check #13,171, D Trump (2 October 2019); #13,955 D Trump (17 October 2019).

140 'Republican senators largely unmoved by Democrats' Trump trial prosecution' *New York Post*, 11 February 2021).

141 Kellner (n 101 above) 77–78.

142 Article of Impeachment against Donald J Trump (2021) by the United States House of Representatives, 117th Congress, 1st Session.

143 Ibid.

Trump ought to be impeached, while in the Senate, seven Republicans voted in favour of the articles. In the Senate that vote fell short of the two-thirds required to convict.[144] Several Republicans who voted in favour of impeachment were called upon to resign. Senate Majority Leader Mitch McConnell delayed the impeachment until Trump was out of office and, while reportedly agreeing that Trump had committed impeachable offences, argued that, as he was out of office, he should not be impeached.[145]

The election succeeded where the US legislative branch did not. The election removed a potential tyrant despite efforts to undermine the results and to dampen turnout. It also succeeded because the judicial branch had not come under the Trump administration's thrall. While the legislative branch continues with investigations into the administration, some members are potentially too reliant on Trump (and Trump not unleashing his supporters upon them) for their own re-election to act against him. Some Republicans whom he has previously viciously attacked now endorse him, including some who became his loudest supporters during the second impeachment.[146] Public humiliation serves Trump well. It was the demos who concluded that his time in office, despite his demands, was at an end.

CONCLUSION: ON THE ROAD TO TYRANNY?

From a gender, poverty and race perspective, there are many in the USA who argue that they live in a tyranny.[147] However, if the benchmark for tyranny is the operation of the US Constitution, despite ever-increasing pressure, tyranny is yet to emerge. Trump is no longer in office, although he may be again in the future. What is more critical than his incumbency of high office is what Trump suggests about the US political and legal order. Just as Julius Caesar did not transform the office of dictator, nor did Trump transform the Office of the President

144 Washington, US Capitol Room H154 (13 January 2021) 'Roll Call 17, Bill Number: H Res 24, 117th Congress, 1st Session'; Office of the Clerk, US House of Representatives: archived from the original on 13 January 2021, retrieved 13 January 2021.

145 Liz Cheney says she won't resign after Wyoming GOP calls for her to step down: archived 9 February 2021, at the Wayback Machine, *Washington Times*, Valerie Richardson, 7 February 2021, retrieved 7 February 2021; Jonathan Martin and Maggie Haberman, 'McConnell is said to be pleased about impeachment, believing it will be easier to purge Trump from the GOP' (*New York Times*, 12 January 2021): archived from the original on 12 January 2021, retrieved 12 January 2021.

146 Kellner (n 101 above) 75–77; A Ware, 'Donald Trump's hijacking of the Republican Party in historical perspective' (2016) 87 Political Quarterly 406.

147 The article has not touched on US foreign relations.

of the USA. The Office of Roman dictator's demise into tyranny began with Sulla and was compounded by Augustus, after whose death the Roman Republic's structures could not recover. In January 2021, the presidency transitioned to Biden – but with accompanying violence and in the face of an ever-growing proportion of Americans who now believe their election system is fraudulent. The real question is whether Trump fatally undermined the US system. Trump may be re-elected. The litmus test for tyranny will be whether he continues in the same vein and, if so, whether he would continue to enjoy support from enough of the Republican Party to make a third impeachment impossible. Could a third impeachment see him removed, and if he continued in the same vein would the continued support from some parts of the Republican Party make a third impeachment impossible? If Trump were to take steps beyond what occurred under impeachment one or impeachment two, is there a real possibility that a third impeachment would succeed in removing him or has the US system become so unravelled that it would be unable to operationalise its key safeguard of removal? If not Trump himself, could a prospective tyrant be elected in his model? Potentially, is there also a future presidential candidate cloaked in performed legitimacy as the anti-Trump who could then take advantage of the denuded political sphere to install tyranny?

In truth, the most important factor in the potential emergence of (post-)Trump tyranny will be the denuding of the public space and contestation. At the centre of this are the two realities: Trump's and the contested space of public debate. A historical denudation of the latter impacted upon the ability to sustain the former and Trump used the denudation and pushed it even further.[148] Within the public sphere of contestation, the role of social media in creating and sustaining Trump's reality is important. Looking forward, it seems the behaviour and regulation of social media companies will play a critical role in creating and sustaining alternative realities and in the fostering of widespread conspiracy theories. Now somewhat chastened by events and with several ultimately banning Trump from their platforms, should Trump run for President again, how would social media companies react? Facebook's appearances before Congress and its adoption of a human rights policy are pre-emptive attempts to stave off regulation.[149] Facebook admits that it is able to impact on voter

148 T Lynch, 'President Donald Trump: a case study of spectacular power' (2017) 88 Political Quarterly 612.

149 Neema Hakim, 'Do not trust Facebook to enforce human rights' (*OpinoJuris*, 22 March 2021); Mayank Aggarwal, 'Facebook "behaving like a North Korean dictator" in Australia' (*The Independent*, 19 February 2021).

turnout and that is an awesome power to leave entirely unregulated.[150] Voting turnout for the last two US presidential elections was very high, but voter suppression is also high. In Georgia, subject of the second impeachment article, a raft of new laws will make it particularly difficult for African Americans to vote.[151] In Georgia, where Trump attempted to interfere in the counting process and claims of rigging unleashed such fury that an election worker had to go into hiding, this is particularly concerning.[152]

Such laws may end up in federal courts, and a further test of the emergence of tyranny will be whether Trump has succeeded in stacking these courts so that they enable such attempts, even if they do not find in his favour in spurious voter fraud petitions relating to the 2020 election. Here, again, the groundwork for Trump was already there, including in the way in which law had evolved to not reign in certain activities like campaign finance, hate speech, institutional violence and inequality, but Trump's presidency exacerbated that process. Law rarely stopped Trump, and it certainly failed to foreclose on his potential future return. Public contestation and the public sphere did more to reign Trump in than law. Nonetheless, the unrelenting denudation, the inability to discern facts amongst the conspiracies and alternative Trumpian reality makes the system ever more precarious. It is arguable that if Pence had failed to certify Biden's election Trump would still be President. That this is contestable shows the extent of the problem.

When examining the events of the past few years, it is important to not get lost in the minutia but to see the whole: we cannot focus only on Trump himself but must see him as part of a broader political and legal context: thus, not just focusing only on Trump but also on the broader legal and political context in which he operates – to consider what aspects enabled him, what aspects of law stopped him, and to contemplate what law's limitations in that context turned out to be. Law is important, but law on its own is not a sufficient buttress against tyranny. Indeed, as outlined above, law and constitutionalism can be enablers of tyranny. A series of questions need to be considered. If Trump had succeeded in stopping the count, would commentators begin to call him an authoritarian constitutional figure, and if not, why not? Why is that a difficult consideration but not for Hungary, Poland, India, Turkey, Russia and others. The answer probably lies in the work of Edward Said and Franz Fanon and others and the use

150 Alex Hern and Julie Carrie Wong, 'Facebook plans voter turnout push – but will not bar false claims from Trump' (*The Guardian*, 17 June 2020).

151 'Georgia's new voting law triggers legal challenges' (*The Economist*, 3 April 2021).

152 Honig (n 7 above) 2.

of knowledge (here of constitutionalism) to describe what is going on elsewhere but to not see it at home. The surprise that greeted Trump's election, and that he withstood two impeachments and might very well return, is only a shock if American (constitutional) exceptionalism is accepted as true. Thick democracy and contestation are absolutely necessary; liberal constitutionalism does not suffice. The lesson here is not that democracy is bad, or that Garett Jones is correct that we should have 10 per cent less of it. The lesson is that taking the health of a democracy for granted and denuding it for short-term gain has long-term consequences.

To finish on a uchronic anecdote: what would have happened if the riot had occurred in December or if Mitch McConnell had not delayed the impeachment process? If Trump were still President, would the Republican Senate have been more or less likely to impeach him? Or another uchronic point: would a secret ballot have altered the outcome? What role did fear, fear of a Trump supporter backlash perhaps, have in the way Senators voted? There is some salve in the idea that Trump did not turn the entire Republican establishment to his needs, but he turned enough of them. James Baldwin, whose assiduous commentary on the negative aspects of the US system remains prescient, stated that '[a] civilization is not destroyed by wicked people; it is not necessary that people be wicked but only that they be spineless'.[153] Trump did not bring about US tyranny, but the acquiescence of those around him just might. There are many limits to liberal constitutionalism, and one of them is that if tyrannical intent exists and is beneficial to enough of the elite, liberal constitutionalism will be of little use as a bulwark in the absence of an active political space.

153 J Baldwin, *The Fire Next Time* (Penguin 1990 [1963]) 52.

Northern Ireland
Legal Quarterly
Autumn Vol. 72 No. 3 (2021): 510–541
Article DOI: 10.53386/nilq.v72i3.979

The myth of associative discrimination and the Court of Justice's great vanishing act: part 2*

Michael Connolly

Portsmouth University
Correspondence email: michael.connolly@port.ac.uk

ABSTRACT

This article complements an article (part 1) recently published in this journal (72(1) NILQ 29–60) contending that the notion of associative discrimination as a term of art renders it so vulnerable to manipulation that it can be used to narrow the scope of the legislation. That argument was rooted in the UK Supreme Court's reasoning in *Lee v Ashers Bakery* [2018] UKSC 49. Part 2 continues the theme, but this time to show that the vulnerability can work the other way, producing, first, an 'extended' notion of associative discrimination and, second, radically broad notions of direct and indirect discrimination. This limb of the thesis also argues that a case heralded as one of associative discrimination, *CHEZ* [2016] CMLR 14, was no such thing. It concludes that the ambitious approach of the European Court of Justice and its Advocates General will blur the traditional form-based distinction between direct and indirect discrimination.

Keywords: extended direct and indirect associative discrimination; *CHEZ*.

INTRODUCTION

This article complements an article previously published in this journal contending that the notion of associative discrimination as a term of art renders it so vulnerable to manipulation that it can be used to *narrow* the scope of the legislation.[1] That argument was rooted in the United Kingdom (UK) Supreme Court's reasoning in *Lee v Ashers Bakery*.[2] The theme continues here, but this time to show that the vulnerability can work the other way, producing, first, an 'extended' notion of associative discrimination and, second, radically broad notions of direct and indirect discrimination. This limb of the thesis also argues that a case heralded as one of associative discrimination

* First published in NILQ 72.AD1 (2021) ADVANCE 1 July/August 1–32.

1 M Connolly 'The "associative" discrimination fiction: part 1' (2021) 72(1) Northern Ireland Legal Quarterly 29–60.

2 [2018] UKSC 49.

was no such thing. All three propositions are rooted in the reasoning of the European Court of Justice (ECJ) in the case of *CHEZ*.[3]

The starting point is a rehearsal of two 'simple' examples. A white worker is dismissed from her job because she married a black man,[4] or a bar denies service to a white woman because she is accompanied by a black man.[5] Although atypical, these scenarios represent what has become known as associative discrimination. In such circumstances, the white person can sue for direct racial discrimination. This potential applies to all three of the principal discrimination Directives of the European Union (EU), covering race,[6] sex and gender reassignment,[7] disability, age, religion or belief, and sexual orientation.[8] Nevertheless, the reason that the white woman can sue is not her association with a black person, but more simply that the treatment was because of race. So, where, for instance, a white manager is dismissed for defying an order to bar black youngsters, he is dismissed because of race. No 'association' is required for liability.[9] To hold otherwise is to assert a myth. The myth is better appreciated when the associative notion is presented with more complex scenarios, such as the one arising in *CHEZ*:

> In a predominantly Roma district, an electricity supplier hostile to Roma people moved meters so high that they could not be read, inconveniencing both Roma and non-Roma residents.

This scenario is far removed from the 'simple' examples, as here a non-Roma victim is 'associated' by the happenstance of the protected characteristic of her neighbours. This state of affairs was characterised neatly by Advocate General Kokott as a matter of 'collateral damage'.[10] As such, she advised that the non-Roma could sue for associative *indirect* (racial) discrimination. The subsequent ECJ decision added that the treatment could amount to *direct* discrimination against the non-Roma, without a mention of it being 'associative'. Whatever the

---

3 Case C-83/14 *CHEZ Razpredelenie Bulgaria AD v Komisia za zashtita ot diskriminatsia* [2016] CMLR 14.

4 See eg Lord Simon, *obiter*, *Race Relations Board v Applin* [1975] 2 AC 259 (HL) 289–290 (on the premise that foster parents discriminated against a local authority for refusing to foster children of colour).

5 See the suggestion by James Comyn QC, approved by Lord Denning MR: *Applin v Race Relations Board* [1973] QB 815 (CA) 828, and 831 (Stephenson LJ), affirmed [1975] 2 AC 259 (HL).

6 Race Directive 2000/43/EC.

7 'Recast' Directive 2006/54/EC.

8 'Framework' Directive 2000/78/EC.

9 *Showboat Entertainment Centre v Owens* [1984] ICR 65 (EAT). See below, text to n 102.

10 Case C-83/14 *CHEZ* [2016] CMLR 14, Opinion of AG Kokott, para AG58.

differences of language and reasoning deployed, the case has been heralded erroneously as an example of associative discrimination.[11]

As with *Lee v Ashers*, an analysis of *CHEZ* shows that recourse to the legislative provisions would have produced simpler questions to ask. In *Lee v Ashers* (where a bakery refused to ice a cake with the message 'support gay marriage'), it would have been whether the refusal was less favourable treatment because of sexual orientation. Instead, the Supreme Court confined its thinking to 'associative discrimination'[12] and thus devised a convoluted and vague associative 'closeness' test. This resulted in a finding of no discrimination because any 'association' with homosexual persons was not 'close enough' to those supporters of same-sex marriage, such as 'parents, the families and friends of gay people' generally.[13] In *CHEZ*, the question should have been the *locus standi* (standing to sue) of the non-Roma victim, a matter for the legislative enforcement provisions, rather than one of substantive law. Instead, the Advocate General produced an extended notion of 'indirect associative discrimination'. As with the Supreme Court, the core error was treating the notion of associative discrimination as a term of art, around which a novel (but broader) version of discrimination was devised. The ECJ did not deploy this terminology, instead producing overbroad, incomplete, and unbounded models of direct and indirect discrimination. Under these, any notion of associative discrimination, along with the conventional boundaries of the direct/indirect framework, disappeared in a great vanishing act. Accordingly, in continuing the contention that associative discrimination is not a term of art, this article highlights the likely missteps when treated as such, and the myth that *CHEZ* was in fact such a case.

Ahead is an appreciation of the governing legislative regime in the context of notions of associative discrimination. This helps inform the subsequent missteps. The substance of the discussion concerns the

11 *Hainsworth v Ministry of Defence* (Leave to Appeal refused, with details) UKSC 2014/0164 (Lord Wilson, Lady Hale, Lord Clarke, Lord Hughes, Lord Hodge) [7]–[9], see also Permission to Appeal results December 2015; *Chief Constable of Norfolk Constabulary v Coffey* [2018] ICR 812 (EAT) [49] (Judge David Richardson). For commentary, see eg *Harvey on Industrial Relations,* Part L, Equal Opportunities, 3(2)(f) [284.01] and (more cautiously for indirect discrimination) 3(3)(a) [291.01]; M Malone, 'The concept of indirect discrimination by association: too late for the UK?' (2017) Industrial Law Journal 46(1) 144; D Mitchell, 'Collateral damage' (2016) 166 (7686) New Law Journal 8-9; M Rubenstein, 'Highlights' 2015 (Sep) [2015] IRLR 746; Á Oliveira, Sarah-Jane King, 'A good chess opening: Luxembourg's first Roma case consolidates its role as a fundamental rights court' (2016) 41(6) European Law Review 865.

12 [2018] UKSC 49 (Lady Hale) [34]: 'This was a case of associative discrimination or it was nothing.'

13 Ibid [33].

Advocate General's Opinion and the court's judgment in *CHEZ*. This begins with the Advocate General's models of associative direct and indirect discrimination, and continues with the court's radical new discrimination models. Within this, there is the court's unorthodox comparison, an ambiguous 'grounds of' approach (ranging from 'related to' to 'hostile intent'), the erosion of form-based indirect discrimination, and a radical extended model of indirect discrimination. There is also a consideration of the edict from *Coleman v Attridge Law*,[14] adopted by the court, and its relationship with the UK harassment case *English v Sanderson Blinds*.[15] Finally, the article identifies the missteps in *CHEZ* and why the case is wrongly regarded as one of associative discrimination (the 'associative myth').

A note of caution. The case of *CHEZ* contains many unconnected strands and incomplete notions, so it is not the easiest to digest. This might be down to the Reference, posing for the court 'no less than ten extremely detailed' questions.[16] These included the meanings of 'comparable situation' and 'apparently neutral practice'.[17] Thus, the court was charged not only with providing general principles of direct and indirect discrimination, but fleshing them out in some detail. Accordingly, some of the analysis on these questions is respectively similarly doctrinal.

ASSOCIATIVE DIRECT DISCRIMINATION AND THE LEGISLATION

Hitherto, associative discrimination has only been considered in the context of *direct* discrimination. Various theories have been advanced in support of making associative discrimination unlawful. Some focus on the third party, or 'associated', member of a suspect class, with a concern over the harm[18] or indignity[19] they suffer via the treatment of someone with whom they are associated. Thus, the black husband of the white worker suffers harm or indignity when his wife is dismissed because of *his* colour. It will become apparent that the legislation suggests a reach further than this.

---

14 Case C-303/06 *Coleman v Attridge Law* [2008] 3 CMLR 27.

15 [2009] ICR 543 (CA) [39] (Sedley LJ). See further Connolly (n 1 above) text to n 26.

16 Case C-83/14 *CHEZ*, Opinion of AG Kokott, para 30 (see also para 37 of the judgment).

17 Ibid, respectively, Questions 2, 6.

18 See eg V Schwartz, 'Title VII: a shift from sex to relationships' (2012) 35 Harvard Journal of Law and Gender 209. Discussed in Connolly (n 1 above) 30–38.

19 Ibid. See also, Case C-303/06 *Coleman v Attridge Law* [2008] 3 CMLR 27, Opinion of AG Maduro, para AG11.

Direct discrimination across the EU equality Directives employs a common formula. For example, the Race Directive provides,

> [D]irect discrimination shall be taken to occur where one person is treated less favourably than another is, has been or would be treated in a comparable situation on grounds of racial or ethnic origin …[20]

This does *not* state 'on grounds of *his* (or *her*) racial or ethnic origin'. The omission of a possessive adjective is key here. Rather than identifying the targeted victim with *a* protected characteristic, the conduct need only be because of a protected characteristic (or suspect class or suspect ground), a somewhat more abstract term. This is in contrast to the United States (US) Civil Rights Act 1964, whose employment discrimination provisions express that the protected characteristic belongs to the victim-claimant.[21]

The Directive's formula requires treatment 'on grounds of' the protected characteristic in question. The 'less favourable' element involves a comparison of how a person would be treated in a comparable situation. The usual approach here is to deduct just the racial (or other suspect ground) element from the comparable situation. Note that the phrase 'or would be treated' allows for a hypothetical comparison. In the simple examples (outlined above), the conduct was on the ground of the race of the third party, or perhaps, the interracial relationship. Either way, if the claimant-victim would not have been rejected in a comparable situation (where the husband or companion were white), the treatment was less favourable.

Thus, the key to the Directives' formulas encompassing associative discrimination is the absence of any requirement that the victim-claimant holds the relevant protected characteristic. Under the US model, even a liberal or purposive interpretation must ultimately refer to the plaintiff's protected characteristic. For instance, in *Tetro v Elliot Popham Pontiac*,[22] a Court of Appeals found that it was discriminatory to dismiss a white worker because his child was mixed race. But the reasoning came back to the statutory formula: '[A] white employee who is discharged because his child is biracial is discriminated against on the basis of his race, even though the root animus for the discrimination is a prejudice against the biracial child.'[23]

---

20 Race Directive 2000/43/EC, art 2(2)(a). The UK legislation is similarly formulated: eg EA 2010, s 13(1); Fair Employment and Treatment (Northern Ireland) Order (FETO(NI)) 1998, SR 1998/3162, Art 3(2A).

21 Civil Rights Act 1964, Title VII, s 706 (42 USC s 2000e-2): 'It shall be an unlawful employment practice for an employer … to discriminate against any individual … because of such individual's race, color, religion, sex, or national origin …'.

22 173 F 3d 988 (6th Cir 1999).

23 Ibid 994.

The Directive's formula has potential to go much further than necessary for this or the 'simple' examples of associative discrimination, which *could* be explained by the targeted victim's own race (a *white* person being associated with a black husband or companion). For instance, a white employee may be less favourably treated by being ordered to bar black guests, or to make the premises more attractive to heterosexuals.[24] This broader legislative intent is confirmed by the inclusion of exceptions expressly confined to the targeted victim's *own* protected characteristic.[25]

However, ECJ authority prior to *CHEZ* seems to extend the Directive's formula even further. In *Coleman v Attridge Law* (where a worker was treated less favourably because of her child's disability), Advocate General Maduro wrote,

> The distinguishing feature of direct discrimination and harassment is that they bear a necessary relationship to a particular suspect classification. The discriminator relies on a suspect classification in order to act in a certain way. ... An employer's reliance on those suspect grounds is seen by the Community legal order as an evil which must be eradicated. Therefore, the Directive prohibits the use of those classifications as grounds upon which an employer's reasoning may be based.[26]

The court endorsed this sentiment with a pithy edict: 'The principle of equal treatment enshrined in the Directive in that area applies not to a particular category of person but by reference to the grounds mentioned in Art.1.'[27] This characterisation of direct discrimination shifts the focus away from the identity of the victim, let alone *anyone* with whom they may be associated. Indeed, whether anyone in the scenario has a protected characteristic seems barely relevant. The 'evil which must be eradicated' is conduct informed by a protected

---

24 Respectively, Race Directive 2000/43/EC, art 2(2) and (in the UK, *Showboat Entertainment Centre v Owens* [1984] ICR 65 (EAT) (see, text to n 102); *Lisboa v Realpubs* [2011] Eq LR 267 (EAT).

25 Eg pregnancy and maternity (Recast 2006/54/EC, art 2(2)(c), referring to 92/85/EEC, art 2). Religious organisations can recruit according to the *victim-claimant's* religion (Framework 2000/78/EC, art 4(2). For the UK, see eg pregnancy and maternity (EA 2010, ss 17, 18), being married or in a civil partnership (s 13(4)). Religious organisations can discriminate because of the victim-claimant's sexual orientation or religion, in the fields of services, public functions, associations, and premises (sch 23, para 2). For the extensive employment exceptions, see sch 9, and for services, sch 3.

26 Case C-303/06 *Coleman v Attridge Law* [2008] 3 CMLR 27, Opinion of AG Maduro, para AG19.

27 Ibid, para 38 (and 50). The 'grounds' alluded to were sexual orientation, religion or belief, disability and age ('Framework' Directive 2000/78/EC, art 1). The principle was applied to the Race Directive 2000/43/EC in Case C-83/14 *CHEZ* [2016] CMLR 14, para 56.

characteristic (race, sex, disability, sexual orientation, etc) rather than any harm or indignity caused to persons belonging to a suspect class. All that is required is conduct of a discriminatory nature and *a* victim. As such, this *Coleman* edict could be labelled 'discriminatory conduct *per se*', or 'discrimination *per se*', or just the '*per s*e edict'.

Thus, in arguing that associative discrimination is not a term of art, the starting point is that the legislative text does nothing to encourage it. A good reason for this is the risk of missteps. Those made by the UK Supreme Court have been highlighted elsewhere.[28] Those made in the other leading case on the matter, *CHEZ*, are considered next.

'EXTENDED' ASSOCIATIVE DISCRIMINATION AND THE *COLEMAN* EDICT: THE CASE OF *CHEZ*[29]

As noted above, in this case, the Bulgarian electricity supplier raised its (outdoor) meters in a predominantly Roma district to at least six metres. This was to prevent tampering.[30] A non-Roma resident, suffering a similar inconvenience, offence and stigma[31] as her Roma neighbours, brought a claim of direct and indirect racial discrimination. This case resembles the 'associative' examples discussed so far, in that the victim-claimant did not belong to the relevant suspect class and (it was assumed) that the treatment was informed by the protected characteristic (race) of others. It differs because these others, the 'third parties', were likewise targeted.

Advocate General's Opinion

In her Opinion for the court, Advocate General Kokott considered that a 'personal link' was not the only conceivable requirement for associative discrimination.[32] She applied her associative theory to direct *and* indirect discrimination.[33]

For direct discrimination, AG Kokott advised that it extended beyond *Coleman* (and the personal link between mother and baby). She gave an example of a group of people refused a table in a restaurant because one of them was black. Here, each of the rejected group's white guests could sue for direct associative discrimination. It made no difference that there may not have been a personal link with the black

28 See Connolly (n 1 above) 30–38, discussing *Lee v Ashers Bakery* [2018] UKSC 49.

29 Case C-83/14 *CHEZ Razpredelenie Bulgaria AD v Komisia za zashtita ot diskriminatsia* [2016] CMLR 14.

30 Ibid para 106.

31 Ibid para 87.

32 Ibid Opinion of AG Kokott, para AG58.

33 Ibid paras AG103–AG109.

guest, say, if they were meeting for the first time.[34] Hence, associative discrimination may,

> be inherent in the measure itself, in particular where that measure is liable, because of its wholesale and collective character, to affect not only the person possessing one of the [protected] characteristics … but also—as a kind of 'collateral damage'—includes other persons.[35]

From this, it would seem that AG Kokott's theory applies to anyone suffering 'collateral damage' from a discriminatory act, and that these persons have suffered direct associative discrimination. This is not as far-reaching as it may first appear. In applying an orthodox comparison, the Advocate General found that, as the Roma and non-Roma residents had been treated equally (badly), any discrimination could not be direct.[36]

According to the Advocate General, if the 'collective' treatment is facially neutral, then it cannot be direct discrimination, associative or otherwise. This raises the question of how far this theory actually extends the notion of direct associative discrimination. AG Kokott's restaurant example shows that where the treatment is facially discriminatory, then there is a case even without a 'personal link'. This may extend beyond the close personal relationship in *Coleman*, but not beyond long-established simple examples. It is no different from the example of the white woman refused admission because she is accompanied by a black man. Whether they had a personal relationship is irrelevant. (They may have been meeting on a blind date.) This example, from the English Court of Appeal, dates to 1973.[37] The Advocate General gave no examples beyond this. As such, it is difficult to conclude that she intended to create a far-reaching theory of direct associative discrimination. On that basis, AG Kokott's theory could be said to require facially discriminatory treatment, applied 'collectively' to a mixed group causing harm to both those with, and those without, the relevant protected characteristic. As such, this theory adds nothing new to the UK legal lexicon and merely confirms that *Coleman* is not confined to 'close personal relationships'. Hemmed in by the orthodox comparison, the theory could not apply to the unusual facts of the case in hand.

Unaware of, or undaunted by, such niceties, from her finding of facially neutral treatment, AG Kokott's analysis defaulted to that of indirect discrimination.[38] Here, her theory came to life. For her, the

---

34 Ibid para AG 59.

35 Ibid.

36 Ibid paras AG85–AG87.

37 See n 5 above.

38 Cf *Lee v Ashers*, ceasing further analysis upon such a finding. See [2018] UKSC 49 [21] and for comment Connolly (n 1 above) 56, 'Treatment "applying to all"'.

facts of *CHEZ* presented a *prima facie* case of associative *indirect* discrimination (with a heavy hint that it could not be objectively justified[39]). In support of her theory here, the Advocate General deployed an example of an employer providing nursery care for children of its full-time employees only. Assuming that the full-timers were predominantly male and part-timers predominantly female, this raised a case of *indirect* sex discrimination. Her associative theory extended the employer's liability towards the part-timers' children, who had likewise suffered.[40] Note here, despite her observation elsewhere in her Opinion that the victims in *CHEZ* had been affected in the same way,[41] this was not laid down as a boundary for her associative theory. Accordingly, it did not matter that the children would not have been affected in the same way as their mothers. This suggests it is not necessary for victims of collateral damage to have suffered the *same* harm as the primary victims. Given this example, her Opinion on the case, and the absence of any boundaries, this associative theory, when applied to *indirect* discrimination, has exceptional potential. In *CHEZ*, there was an association merely by the happenstance of the protected characteristic of the claimant's neighbours. But the principle espoused here applies to anyone with the misfortune to have been harmed by a practice adversely affecting a suspect class. These victims need not be neighbours in the geographical, or any, sense, save for being harmed by the same practice. AG Kokott's associative theory is more about associated *harm* than anything else, suggesting that *anyone* harmed by a discriminatory act has suffered discrimination. (The ramifications are considered, further below, in the discussion on the court's finding on indirect discrimination.)[42]

For AG Kokott's associative theories, everything turns on the comparison. It would seem to have much more potential for indirect discrimination than for direct discrimination. However, her efforts failed to influence the court, as her associative theories vanished as the court delivered its own radical models of discrimination.

---

39 Case C-83/14 *CHEZ*, Opinion of AG Kokott, AG106-AG109 and (for justification) para AG139.

40 Ibid para AG107.

41 Elsewhere in her Opinion, she said that Roma and non-Roma were affected in the same way, but this was not laid down as a boundary for her theory on direct associative discrimination: para AG98.

42 See 25–27.

The judgment

Direct discrimination

For all the talk of associative discrimination by its Advocate General, the court's judgment did not mention it when holding that the claimant could sue for *direct* discrimination, or in the alternative, for indirect discrimination.[43] After iterating that the equality legislation should not be given a restrictive interpretation,[44] the court repeated the edict from *Coleman*, that the legislation is to combat discrimination *per se* and 'not to a particular category of person'.[45] From there, the court produced a judgment even more radical than its Advocate General's Opinion. It came to these findings via an unorthodox approach to the 'less favourable' element, an ambiguous presentation of the 'grounds of' question, and indications that a case could turn on motive rather than form, which returns the matter to the *Coleman* edict. These features are discussed next, in turn.

The comparison for the less favourable element

AG Kokott's contrary Opinion relied on an orthodox comparison, incorporating all relevant facts save for the protected characteristic in question. For the court, however, rather than comparing the treatment of different racial groups, the comparison was between those whose meters had been raised, and those for whom they had not.[46] This departure from orthodoxy had more to do with meters and place of residence than with ethnicity. It merely compared those who had been mistreated with those who had not. Such a question will return the same answer every time. It proves no more than that the defendant treated one *district* (rather than ethnic group) less favourably than it did another; if the protected characteristic in question were race, this is an 'apparently neutral practice', and so apt for an indirect discrimination analysis.[47]

---

43 Case C-83/14 *CHEZ*, para 50.

44 Ibid para 42, citing Case C-391/09 *Runevic-Vardyn v Vilniaus Miesto Savivaldybes Administracija* [2011] 3 CMLR 13, para 43.

45 Ibid para 56. In Case C-303/06 *Coleman v Attridge Law* [2008] 3 CMLR 27, para 38 (and 50); at para 64(1) the court ruled simply that the legislation was 'not limited only to people who are themselves disabled'. The Opinion advocating an 'associative' theory was written by AG Maduro [2008] 3 CMLR. 27, paras AG9–AG14 and AG19. See further Connolly (n 1 above) 32–36.

46 Ibid para 90.

47 The Bulgarian court had held that this was direct discrimination on the ground of 'personal situation', a protected characteristic under Bulgarian law: ibid paras 13 and 26.

EU legislation requires a comparison for direct discrimination as a matter of substantive law, although it is arguable that it affords some discretion in fashioning a 'comparable situation'. Less discretion is afforded by the UK legislation, requiring 'no material difference between the circumstances relating to each case'.[48] Again, this is an element of the direct discrimination formula. That said, UK case law has shown a willingness to dispense with the comparison, but only where a comparison would prove problematic and the 'reason why' was established.[49] Otherwise, the comparison is 'compulsory'.[50] Of course, making an orthodox comparison in *CHEZ* was quite unproblematic, as the Advocate General demonstrated.

When employed, orthodox comparisons help distinguish the protected characteristic in question, and thus isolate direct discrimination from merely unfair, or equally bad, treatment (the latter suggesting an indirect discrimination analysis is due, as the Advocate General deduced). The CHEZ comparison could not do this. The court did not explain this departure from orthodoxy or from its Advocate General's Opinion. Given that the Reference expressly requested clarity on the meaning of the comparison,[51] this unorthodoxy is all the more puzzling. Two explanations are ventured here.

First, although not expressed as such, the court may have been led into this by some notion of associative discrimination. The comparison made in *CHEZ* distinguished the treatment of one group (targeted Roma and non-Roma who were associated by neighbourhood) from that of another (those not targeted and associated by neighbourhood). Second, a more likely, or perhaps complementary, explanation rests in the *Coleman per se* edict, which, unlike AG Kokott's associative theory, *was* cited. With its focus on the defendant's conduct (and presumably consequent harm), rather than the protected characteristic of the victim-claimant, this edict absorbs the non-Roma as 'primary victims'. This explanation is also supported by the subsequent reasoning, which duly focused on the defendant's stereotyping of the Roma, which could be interpreted as the critical factor in the finding of direct discrimination.

A relaxed comparison facilitates this approach, which in effect has the potential to convert form-based indirect discrimination into direct discrimination. Indeed, given the novel consequence, and that an apparently neutral practice can be converted into direct discrimination

48 EA 2010, s 23; FETO(NI), art 3(3).

49 *Shamoon v Chief Constable of the RUC* [2003] ICR 337 (HL) [11] (Lord Nicholls), applied, *Chief Constable of Norfolk v Coffey* [2019] EWCA Civ 1061 [76]. See further below, 'The *Coleman* edict and *Sanderson*-type Cases', 18.

50 *Glasgow CC v Zafar* [1998] ICR 120 (HL).

51 Case C-83/14 *CHEZ*, Question 2: para 37; Opinion of AG Kokott, para 30.

upon the 'grounds of' element, then the precise meaning of that element needs to be defined. This is especially so because the Reference expressly required a precise definition of direct discrimination.[52] The judgment fell short here.

'Grounds of'– reasons relating to, or motivated by, discrimination?

It was clear that the judgment considered that a defendant's discriminatory reasoning was relevant to the 'grounds of' question, but it clarified neither its precise meaning nor role. It appeared to pitch two models either side of the conventional 'grounds of' approach.

The court committed considerable attention to evidence implicating the supplier as acting on Roma stereotyping or prejudice. This, aided by the relaxed comparison, presented a presumption of direct discrimination (for the defendant to rebut 'exclusively on objective factors unrelated to any discrimination on the grounds of racial or ethnic origin').[53] Later in the same section of the judgment, the court wrote that there is direct discrimination where the conduct was 'introduced and/or maintained for reasons relating to the ethnic origin common to most of the inhabitants of the district concerned …'[54] A reader predisposed to the conventional approach to the 'grounds of' question, might benignly suppose that this section is following that conventional approach, despite the imprecise language employed. After all, nothing in this section of the judgment expressly declared a major change of direction. However, the danger with such imprecision is that it opens the judgment to different interpretations, especially given the element's pivotal role in the novel suggestion that an apparently neutral practice can amount to direct discrimination. A less forgiving reader of this section could detect two alternative thresholds for the 'grounds of'' question: either 'reasons relating to' race, or (racial) 'stereotyping or prejudice'.

A rubric, 'reasons *relating to*' race, is broader than 'on grounds of' race (although not as broad as the harassment provisions requiring only '*conduct* relating to' a suspect class).[55] But it is broad enough, especially given the relaxed comparison, to encompass many cases of facially neural treatment, ordinarily treated as indirect discrimination. This could prove critical. A natural application of the phrase would

52 Ibid, Questions 2, 3, 4: para 37; Opinion of AG Kokott, para 30.

53 Ibid paras 81–84.

54 Case C-83/14 *CHEZ*, para 91.

55 Race Directive 200/43/EC, art 2(3); Recast 2006/54/EC, art 2(1)(c); Framework 2000/78/EC, art 2(3).

capture the likes of *Bressol*,[56] for example, where the ECJ accepted that restrictions amounted only to *indirect* discrimination, despite being drafted in barely disguised discriminatory terms (a residence requirement is typical), with the intention to restrict other EU nationals from utilising the host nation's education benefits. These restrictions were imposed for reasons 'relating to' nationality. It was unlikely that stereotyping or prejudice could convert them into direct discrimination, but the 'relating to' rubric surely could. In the UK, in *Orphanos v Queen Mary College*,[57] to avoid (higher) overseas fees, a requirement was imposed on students to be ordinarily resident within the European Community (EC) for three years. It was imposed knowingly against non-EC nationals with the immediate goal of curtailing public expenditure on education. The House of Lords treated the requirement as indirect discrimination, with the consequence that the claimant was awarded no compensation.[58] Yet, the requirement clearly was imposed for a reason relating to nationality, as the House of Lords found when rejecting the college's justification defence.[59] It would seem that a post-*CHEZ* claimant could have been awarded compensation for direct discrimination. This approach could also deprive many defendants of a good defence. In *Greater Glasgow Health Board v Carey*,[60] for example, the employer justified a denial to a health visitor's request to move to part-time work because patients required regular daily personal contact. Given that any employer (especially one with a significant Human Resources Department) is likely to be aware of the impact on women, this refusal, although not on 'grounds of' sex, could

56 Case C-73/08 *Bressol v Gouvernement de la Communauté Franĉaise* [2010] 3 CMLR 559. See also Case C-209/03 *R (Bidar) v Ealing LBC* [2005] ECR I-2119; followed by the UK Supreme Court in *Patmalniece v Secretary of State for Work and Pensions* [2011] UKSC 11. 'Nationality' discrimination is unlawful by Treaty on the Functioning of the European Union, art 18 (ex art 12 Treaty establishing the European Community). For a summary of the political background, see, S Garben, 'Case Comment on *Bressol*' (2010) 47(5) Common Market Law Review 1493, 1496–1498.

57 [1985] AC 761 (HL).

58 Ibid respectively 772–773 and 774–775. Although the fee was not justified, the Pyrrhic victory was because, at the time, for indirect discrimination only, compensation could only be awarded if the requirement was intended to discriminate: Sex Discrimination Act 1975, s 66(3); Race Relations Act 1976, s 57(3). This has since been ameliorated. See *JH Walker v Hussain* [1996] ICR 291 (EAT), at 299–300; *London Underground v Edwards* [1995] ICR 574 (EAT), where an 'awareness' of the discriminatory impact was enough for compensation to be payable. See now EA 2010, ss 119(6), 124(5); FETO(NI), art 39(3).

59 Ibid 773.

60 [1987] IRLR 484 (EAT). The employer justified the refusal by offering five half-days per week, instead of the requested three whole days.

be said to have been made for a reason 'relating to' sex, and as such, under *CHEZ*, the defence would be lost.[61]

These examples merely illustrate how the 'relating to' phrase, once combined with the relaxed comparison, could change the outcome of cases ordinarily treated as indirect discrimination. Of course, such a low threshold for direct discrimination liability does not accord with the legislative phrase 'grounds of' nor the legislative scheme, which expressly provides an objective justification defence only for indirect discrimination, which is expressed as form-based, requiring only *apparently* neutral practices that put suspect classes at a particular disadvantage.[62]

At the other extreme, the court's extensive detailing of the supplier's prejudice and stereotyping suggests that a discriminatory motive is required for direct discrimination liability. This at least could put a check on an otherwise extraordinary reach of this relaxed version of direct discrimination. There is further evidence supporting this interpretation in the subsequent case of *Achbita*,[63] where the court found that an employer's blanket ban on visible signs of religious, philosophical, or political beliefs amounted to indirect discrimination. This was despite the ban obviously 'relating to' religion, notably that of the headscarf-wearing Muslim worker who was dismissed for defying the order. Accordingly, in *Achbita*, AG Kokott distinguished *CHEZ* as turning on the supplier's motive,

> As is clear from the judgment in *CHEZ* ... [82], the Court considers a measure taken on the basis of stereotypes and prejudices in relation to a particular group of individuals to be an indication of direct discrimination (based on ethnic origin).[64]

Even with these indications that the court looks for a discriminatory motive, the next question is what quite this means. Within the notion of motive, there is a range of states of mind that might be required for liability. These could range from malice, hostility, prejudice, stereotyping, or just foresight or even constructive knowledge of the impact of the conduct, such as having ought to have been aware that the district targeted for raised meters was predominantly Roma.

---

61 It could not be argued as genuine occupational requirement (GOR), as there was no requirement for a man to do the job. For GORs see, EA 2010, sch 9, part 1, para 1; Sex Discrimination (NI) Order 1976, SR 1976/1042, art 10; 'Recast' Directive 2006/54/EC 14(2).

62 Race Directive 2000/43/EC, art 2(2)(b); Recast 2006/54/EC, art 2(1)(b); Framework 2000/78/EC, art 2(2)(b). Emphasis supplied.

63 C-157/15 *Achbita v G4S Secure Solutions NV* [2017] 3 CMLR 21.

64 Ibid para AG55, n 30.

Whatever the level of motive envisaged, importing notions of discriminatory intent into direct discrimination without qualification brings with it issues. Many claims would be more difficult to bring, especially if claimants were unduly burdened with proving requirements from higher in the range, such as a malicious intent.[65] It also risks, or encourages, the recognition of 'benign motive' defences, whereby even though there is a discriminatory reason for the treatment, the defendant demonstrates a benign motive, such as customer preference, chivalry, or protection from harassment or even violence.[66] One must presume that this would be against policy[67] or any purpose that could be attributed to equality legislation, but it might reignite long-standing arguments and divisions on the matter.[68]

A shift of emphasis towards a discriminatory motive brings to mind the US jurisprudence. Instead of a form-based approach, the US counterparts to direct and indirect discrimination are known respectively as intentional and non-intentional discrimination.[69] The US courts tend to utilise a comparison as an evidential tool in proving

---

65 S Fredman, *Discrimination Law* 2nd edn (OUP 2011) 203–214.

66 Respectively, *Diaz v Pan Am* 442 F 2d 385 (US, 5th Cir 1971), *certiorari* denied, 404 US 950 (1971) (preference for female cabin crew); *Segor v Goodrich Actuation Systems Ltd* (2012) UKEAT/0145/11/DM (US arms contract stipulated 'no French nationals'); *Hafeez v Richmond School* (Industrial Tribunal, 27 February 1981) (parents' preference for pupils to be taught by 'English teachers'). *Ministry of Defence v Jeremiah* [1980] QB 87 (CA) (women not required to work in dirty part of factory); *Grieg v Community Industry* [1979] ICR 356 (EAT) (woman denied work with all-male decorating team); *Amnesty International v Ahmed* [2009] ICR 1450 (EAT) (Sudanese national rejected as reporter because of risk of violence).

67 In Case C-188/15 *Bougnaoui v Micropole SA* [2017] 3 CMLR 22, para 40, it was held that a customer preference for workers not to wear a headscarf could not amount to a GOR defence. This was because these were 'subjective considerations, such as the willingness of the employer to take account of the particular wishes of the customer' (at [40]).

68 In the UK, on no less than eight occasions, the House of Lords/Supreme Court has entertained the issue and often divided on it: *R v Birmingham CC ex p EOC* [1989] 1 AC 1156; *James v Eastleigh BC* [1990] 2 AC 751; *Nagarajan v LRT* [2000] 1 AC 501; *Chief Constable of West Yorkshire v Khan* [2001] UKHL 48; *Shamoon v Chief Constable of the RUC* [2003] ICR 337; *R (European Roma Rights Centre) v Immigration Officer at Prague Airport* [2005] 2 AC 1 (HL); *St Helens MBC v Derbyshire* [2007] ICR 841; *R (E) v Governing Body of JFS* [2010] 2 AC 278 (SC). In addition, the issues of knowledge of the protected characteristic, and (for two Law Lords) discriminatory intent, arose in *Lewisham LBC v Malcolm* [2008] 1 AC 1399 (HL).

69 For discussions on the precise meaning of intent in the US, see M Selmi, 'Proving intentional discrimination: the reality of Supreme Court rhetoric' (1997) 86 Georgetown Law Journal 279; E Schnapper, 'Two categories of discriminatory intent' (1982) 17 Harvard Civil Rights–Civil Liberties Law Rev 31.

the discriminatory, or 'hostile', intention, rather than as a necessary element.[70] The difference is that the US model exists within the framework of developed jurisprudence, notably its 'pretext' doctrine (discussed below).

This is not to say that stereotyping cannot be evidence that a reason for the treatment is a protected characteristic,[71] notably where the grounds for the treatment are 'not obvious':[72] for example, where a bartender says to a homeless black man dressed in rags, 'I do not serve people like you.'[73] Hitherto, direct discrimination requires no motive, only a discriminatory ground, or cause, for the treatment. The distinction was set out by Lord Nicholls in *Nagarajan*,

> The crucial ['grounds of'] question ... is to be distinguished sharply from a second and different question: if the discriminator treated the complainant less favourably on racial grounds, why did he do so? The latter question is strictly beside the point when deciding whether an act of racial discrimination occurred.[74]

It might be that the *CHEZ* judgment cited stereotyping and prejudice merely as evidence of something else. The difficulty is that the 'something else' was not made clear. The reference to the reason being 'related to' ethnicity suggests such a low threshold, that the evidence of stereotyping or prejudice was unnecessary. All that could be required under this standard would be foresight that the practice would harm those in a predominantly Roma district. Instead of pitching two models either side of the conventional 'grounds of' approach, the matter would have been clearer if it had deployed this legislative format and explained the threshold it applied. The obligation upon the court was all the more so given the request for clarity on the meaning of direct discrimination expressed in the Reference.[75]

---

70 See eg Civil Rights Act 1964, Title VII (employment), s 703 (42 USC s 2000e-2(m)). For the evidential role of the comparison, plaintiffs use a 'similarly situated' rubric. See *International Brotherhood of Teamster v US* (1977, Sup Ct) 431 US 324, n 15: 'Proof of discriminatory motive is critical, although it can in some situations be inferred from the mere fact of differences in treatment.' Cited by the Equal Employment Opportunity Commission Compliance Manual, 604.1(a) .

71 See eg *Alexander v Home Office* [1988] 2 All ER 118 (CA) 120h: 'He displays the usual traits associated with people of his ethnic background being arrogant, suspicious of staff, anti-authority, devious and possessing a very large chip on his shoulder ... that seems too common in most coloured inmates.' This was evidence that a refusal of (preferable) kitchen work for a prisoner was directly discriminatory.

72 *Nagarajan v LRT* [2000] 1 AC 501 (HL) 511.

73 See also *R (E) v Governing Body of JFS* [2010] 2 AC 278 (SC) [21] (Lord Philips).

74 *Nagarajan v LRT* [2000] 1 AC 501 (HL) 511.

75 Case C-83/14 *CHEZ*, Questions 2, 3, 4: para 37; Opinion of AG Kokott, para 30.

Form, motive, pretext and the relationship with indirect discrimination

It was suggested above that the relaxed comparison permitted a conventional case of indirect discrimination to be analysed as direct discrimination. This appeared to turn on the 'grounds of' element, requiring a reason 'related to' race or based on racial 'stereotyping or prejudice'. In addition to providing no precision as to the meaning of this element, the *CHEZ* judgment changed this element's relationship with indirect discrimination.

In addition to the unorthodox comparison, this reasoning appears to depart from the conventional form-based approach, which in this case would be the apparently neutral conduct based on a place of residence. Under a form-based approach, a facially neutral practice cannot be converted into direct discrimination by the reason for the treatment. For indirect discrimination, the reasons for the treatment belong to the objective justification defence, which of course would be greatly undermined by evidence of stereotyping or prejudice.[76] Hence, the dominating 'grounds' feature is at odds with previous ECJ practice, notably in cases of intentional discrimination against non-nationals wanting to exploit a host nation's advantageous benefits,[77] using the simple expedient of drafting 'their way out of direct into indirect discrimination'.[78]

The form-based approach is supported by the Directive (and the UK legislation). The Directive's definition of indirect discrimination, requiring an '*apparently* neutral provision, criterion or practice',[79] suggests that even a deceitful neutral practice should be analysed only as *indirect* discrimination.[80] This accords with previous ECJ practice,

---

76 See eg Case 96/80 *Jenkins v Kingsgate* [1981] ICR 592 [11]: '[The defence should] in no way related to any discrimination based on sex.'; *R v Secretary of State for Employment, ex p EOC* [1995] 1 AC 1 (HL) 30 (Lord Keith): '[A] gross breach of the principle of equal pay … could not be possibly regarded as a suitable means of achieving a [legitimate aim]'. See also, *Orphanos v QMC* [1985] AC 761 (HL) 772–773, and *R (Elias) v Secretary of State for Defence* [1985] AC 761 (HL) [161]–[162].

77 Case C-73/08 *Bressol v Gouvernement de la Communauté Franĉaise* [2010] 3 CMLR 559. See n 56 above and accompanying text.

78 *Patmalniece v Secretary of State for Work and Pensions* [2011] UKSC 11[73] (Lord Walker).

79 Race Directive 2000/43/EC, art 2(2)(b); Recast 2006/54/EC, art 2(1)(b); Framework 2000/78/EC, art 2(2)(b). Emphasis supplied.

80 The phrase has been criticised for allowing (deceitfully) disguised bigotry to go unchecked as direct discrimination. See Frej Klem Thomsen, 'Stealing bread and sleeping beneath bridges — indirect discrimination as disadvantageous equal treatment' (2015) 2(2) Moral Philosophy and Politics 299, 300; S Atrey, 'Redefining frontiers of EU discrimination law' [2017] Public Law 185, 189-190.

but undermines *CHEZ*. Again, this was not addressed by the *CHEZ* court, but given that its decision seems at odds with this legislative definition, *and* that this was a specific question in the Reference,[81] a clarification as to its meaning and relevance surely was required.

The reasoning presents a particular challenge to the UK legislation, which assumes, in its remedies provisions, that indirect discrimination can be intentional *or* unintentional.[82] On this basis, a UK tribunal or court must decide if indirect discrimination is intentional or not. If it were to hold instead that a discriminatory motive converted apparently facially neutral treatment into direct discrimination, it would render the provision redundant.[83]

As noted above, allowing the reason for the conduct to dictate the 'direct or indirect' question chimes with the US approach, which demarcates over intent, rather than form. This exists within a developed jurisprudence, notably its 'pretext' doctrine, where, following the defendant's rebuttal showing a non-discriminatory reason for the treatment, a plaintiff may submit evidence of a discriminatory motive behind this apparently neutral reason, thus exposing it as a pretext for intentional discrimination.[84] The *CHEZ* judgment evokes a US-style model, but without a 'pretext' framework of shifting burdens or a body of jurisprudence.[85]

The *Coleman* edict and *Sanderson*-type cases

CHEZ has implications also for scenarios where nobody's protected characteristic is involved. The most obvious source of legal principle for the decision in *CHEZ* is the *Coleman per se* edict, where all that is required for direct discrimination is conduct of a discriminatory

81 Question 6: Case C-83/14 *CHEZ*, para 37; Opinion of AG Kokott, para 30.

82 See now EA 2010, s 119(6), 124(5); FETO(NI), art 39(3); Sex Discrimination (Northern Ireland) Order 1976 SR, 1976/1042, art 65(1B); Race Relations (Northern Ireland) Order 1997, SR 1997/869, art 54(3); Employment Equality (Sexual Orientation) Regulations (Northern Ireland) 2003 (SORs), SR 2003/497, reg 36(2); Employment Equality (Age) Regulations (Northern Ireland) 2006, SR 2006/261, reg 43(2); Equality Act (Sexual Orientation) Regulations (Northern Ireland) 2006 (SORs), SR 2006/439, reg 36(3). Indirect discrimination is not specified in the Disability Discrimination Act 1995 applying to Northern Ireland.

83 This would breach a 'cardinal rule' of statutory interpretation: *Re Florence Land Co* (1878) 10 Ch D 530 (CA) 544 (James LJ); *Hill v William Hill (Park Lane) Ltd* [1949] AC 530 (HL) 546–547 (Lord Simon).

84 See eg *Domingo v New England Fish Company* 727 F 2d 1429 at 1435–1436 (9th Cir 1984): where an employer ought to be aware that its word-of-mouth recruitment policy (the 'neutral' practice) had a discriminatory effect, it will be liable for intentional direct discrimination, the word-of-mouth recruitment being a mere pretext.

85 The equality Directives allocate just one shift from the claimant's *prima facie* case to the defendant's rebuttal'. See eg Race Directive 2000/43/EC, art 8.

nature and *a* victim. This principle is wider than necessary even for this radical decision. The conduct in *CHEZ*, even if taken as facially neutral, was related to ethnicity, as both the claimant and the Roma were collectively targeted and harmed. Unlike the *CHEZ* scenario, the edict encompasses the situation where *no one* with a protected characteristic is related to the treatment.

This brings to mind another case of extended liability, *English v Sanderson Blinds*,[86] decided by the English Court of Appeal. Here, it may be recalled, the victim was harassed by colleagues using homophobic sexual innuendo in circumstances where the tormentors knew the victim was not gay. (The victim was aware of his tormentors' knowledge.) A majority held that there could be liability here for harassment 'on the grounds of' sexual orientation.[87] This statutory definition was broad enough to encompass this conduct, but a policy concern regarding privacy was evident in the decision. It meant that, in order to complain, anyone harmed by discriminatory harassment was not obliged to reveal whether or not they belonged to the relevant suspect class.[88] Although the decision exploits the potential of the statutory definition (not limited with a possessive adjective, his/her), it does not go beyond it and accords with an important policy consideration of persons being able to combat harassment without having their privacy violated.

A subsequent question hanging over this case is whether it would apply to direct discrimination. After citing the case and some of its majority and dissenting reasoning, the Supreme Court in *Lee v Ashers* offered no opinion on this question.[89]

The rather obvious barrier to the '*Sanderson* principle' migrating to direct discrimination is that, unlike harassment, discrimination requires *less* favourable treatment and the consequent comparison. Where the victim has no relevant protected characteristic, it is not possible to envisage a comparator without the victim's protected characteristic, as a conventional comparison requires. A similar problem arises with direct *perceived* discrimination, where again, victims have no relevant protected characteristic, but this time defendants wrongly think that they do. When presented with this scenario, in *Chief Constable of Norfolk Constabulary v Coffey*, Judge Richardson, sitting in the Employment Appeal Tribunal (EAT), simply asked: how would the defendant have treated a person he did not

86 [2009] ICR 543 (CA). See further Connolly (n 1 above) text to n 26.

87 Employment Equality (Sexual Orientation) Regulations 2003, SI 2003/1660, reg 5. See now EA 2010, s 26, which replaced the term 'grounds of' with 'related to'.

88 [2009] ICR 543 (CA) [37]–[39] (Sedley LJ).

89 [2018] UKSC 49 [30]–[31].

perceive to have the protected characteristic in question?[90] That is linguistically neat, but it cannot disguise that it changes nothing bar the *defendant's* perception. Direct discrimination requires a comparison of how the defendant treated (or would have treated) *others*,[91] not how a different defendant would have treated the same victim. This approach was argued out in a conventional direct discrimination case, *Grieg v Community Industry*.[92] An employer maintained that its refusal to employ a woman on an all-male decorating team should be compared with a refusal to employ a man on an all-female team. The EAT rejected such a comparison because it involved changing the *defendant's* circumstances. The proper comparison should be how the same employer would have treated a different applicant, here a man, applying for the *same* job.[93]

In some cases, it would be possible to change an *attribute* of the comparator. This is where the misperception was triggered by an attribute of the claimant, say, a crucifix, a headscarf, a turban, or less tangible features, such as an effeminate mannerism or an African-sounding name. Here, the comparison could be with how a person without the relevant attribute would have been treated. But, such a list will tail off into such intangible or unspecifiable matters that no defendant could express *why* he or she was mistaken. It might also be that the defendant, for reasons of embarrassment, say, would not admit to what triggered the mistake. Thus, it cannot be said that a meaningful comparison is always possible for perceived discrimination.

The approach in *Coffey* implemented an express legislative policy commitment that perceived discrimination should be actionable.[94] Indeed, the Court of Appeal, while approving of Judge Richardson's test, nevertheless abandoned the comparison altogether, on the basis it was unnecessary where the 'grounds', or 'reason why', question had been answered.[95]

Whatever the merits of distorting the comparison for perceived discrimination, the matter is yet to be addressed with a *Sanderson* situation. One would expect at least a policy imperative in support of this, either legislative, or inferred, as in *Sanderson*. This suggests that

90 [2018] ICR 812 (EAT) [62].

91 A point made with apparent approval, by counsel in the context of the political opinion claim and the (disapproved) notion that the discrimination could be on the ground of the *defendant's* religious/political belief: *Lee v Ashers Bakery* [2018] UKSC 49 [44]–[45].

92 [1979] ICR 356 (EAT).

93 Ibid 360–361.

94 EA 2010, Explanatory Note 63.

95 [2019] UKCA 1061 [76]–[77] (Underhill LJ), citing *obiter dictum* from *Shamoon v Chief Constable of the RUC* [2003] ICR 337 (HL) [11] (Lord Nicholls).

either a relaxed comparison (as in *CHEZ*) or none at all (the *per se* edict) would be required for direct discrimination. Thus, for UK courts, it may take a discrimination case raising a similar policy issue for the question to be faced. It should be noted here that, where the facts satisfy both the definitions of direct discrimination and harassment, the complaint *must* be treated as one of harassment.[96] So, it would take a rare discrimination case to raise a privacy issue. It might be that this arises where the harassment provisions are not available. The EA 2010 and its Northern Ireland counterparts exclude harassment related to religion or belief, or sexual orientation, from the provision of services.[97] Thus, a *Sanderson*-type case on these excluded areas might force the issue. For example, under a new regime wanting to encourage more heterosexual custom,[98] a bar manager may abuse a customer for his effeminate appearance, despite knowing that the customer is not actually homosexual. In *Sanderson*, the treatment was because of some attributes that the tormentor associated with sexual orientation. These were the claimant's attendance at boarding school and place of residence (a well-known centre of gay society). In this scenario, the attribute is the customer's appearance. A comparator without this effeminate appearance may be considered efficacious. But where the manager could not specify or articulate *why* he thought the customer's appearance was effeminate, the only comparator is a different manager. Thus, as a matter of principle, this issue remains to be addressed.

---

96 EA 2010, s 212(1): '"detriment" does not ... include conduct which amounts to harassment'. Thus, if the conduct amounts to harassment, s 212 dictates that there is no 'detriment' for the provisions on employment discrimination (eg EA 2010, s 39). For a similar proviso in Northern Ireland, see FETO(NI) 1998, art 2(2); SORs, reg 2(4); Disability Discrimination Act 1995, s 18D(2); Sex Discrimination (Northern Ireland) Order 1976, SR 1976/1042, art 2(2); Employment Equality (Age) Regulations (Northern Ireland) 2006, SR 2006/261, reg 2(3). The Race Relations (Northern Ireland) Order 1997, SR 1997/869, has no such proviso. Where harassment is excluded, EA 2010, s 212(5) (but not the SORs or FETO) lifts the proviso, allowing 'harassment' claims where the facts satisfy the definition of direct discrimination.

97 EA 2010, s 29(8), or premises s 34(4)); exercise of public functions (ss 28(8), 33(6), 34(4), 35(4)); associations (members or guests s 103(2)); school education (s 85(1), which also excludes harassment related to gender reassignment). See FETO(NI); Equality Act (Sexual Orientation) Regulations (Northern Ireland) 2006, SR 2006/439 (the harassment provision in reg 3 was quashed under judicial review because of an 'absence of proper consultation': *The Christian Institute v The Office of the First Minister and Deputy First Minister* [2007] NIQB 66; [2008] ELR 146 [34] and [43]).

98 See *Lisboa v Realpubs* [2011] Eq LR 267 (EAT), where a new regime ordered to make the premises more attractive to heterosexuals and correspondingly less so to its traditional homosexual customers.

In *Lee v Ashers*, the Supreme Court avoided the matter by confining its thinking to associative discrimination.[99] As such, the bar's customer would be well-advised not to compress a claim into a notion of associative discrimination and its accompanying rule that the association must be 'close enough'. At EU level, the *Coleman* edict, if taken literally, would readily embrace *Sanderson*-type cases, as no comparison seems to be required and, as *CHEZ*, demonstrates, the ECJ will not be confined by any restrictive notion of associative discrimination. There are no similar exclusions of harassment in the equality Directives, but given the referral system, which is less adversarial, should a *Sanderson*-type case arise, a ruling of both discrimination and harassment by the ECJ would be welcome, not least to clarify the reach of the *Coleman* edict, notably with the role and status of both the comparison and 'grounds of' elements.

A further danger of the Coleman edict – 'anti-purpose' claims

The edict given in *Coleman* was unnecessarily wide for the facts of that case. The existence of a third party with a relevant protected characteristic made it unnecessary to isolate the reasoning to the employer's conduct. (The treatment alluded to the victim's disabled child.) This over-breadth in itself was not unusual. As any student of ECJ jurisprudence will know, unlike the common law's preoccupation with *rationes decidendi*, the court is comfortable producing broad principles under which many a case can be solved. This edict is no different, but without boundaries, it carries a danger that became apparent in a series of English cases, none of which were cited in the ECJ.[100]

Working within that edict, the *CHEZ* court found that evidence of the 'relating to' question was the supplier's stereotyping of Roma people, which effectively drew on the supplier's discriminatory motive. As noted above, the court gave no precise meaning here. The edict could be confined narrowly only to where the conduct has a *hostile* motive (such as the supplier's stereotyping of Roma). At the other end of the scale, it could require no more than an awareness of the discriminatory effect of the conduct. It is at this end of the scale that the danger lurks, where the conduct is merely 'relating to' the suspect class.

The danger is this. While it is convenient to fall back on a common edict such as this to explain and embrace notions of perceived, associated, and third-party discrimination, as well as, perhaps, the

99 See Connolly (n 1 above) 36,39,47 and 57.

100 The cases are: *Showboat v Owens* [1984] ICR 65 (EAT); *Wheeler v Leicester Council* [1985] AC 1054 (CA and HL); and *Redfearn v Serco (t/a West Yorkshire Transport Service)* [2006] EWCA 659). Westlaw and Eur-Lex.Europa searches (25 November 2020) revealed that the ECJ has never cited any of these cases.

'*Sanderson* principle',[101] and even *CHEZ*, its rather universal nature could encompass 'anti-purpose' claims. These can arise where a party is sued because of its *anti*-discrimination conduct, it being conduct 'relating to' a protected characteristic. Note here that such conduct could be characterised as being 'on grounds of' a protected characteristic, and so the legislative formulas are susceptible to these anti-purpose claims, as the examples below will demonstrate.

A good starting point for the English series of cases is *Showboat v Owens*,[102] where, it will be recalled, a white manager was dismissed for defying an order to bar black youths. In holding that this amounted to direct discrimination, Browne-Wilkinson J offered an opinion wider than necessary for this decision:

> [T]here seems to be no stopping point short of holding that any discriminatory treatment caused by racial considerations is capable of falling within section 1 of the [Race Relations] Act of 1976.[103]

The substance of this statement is strikingly similar to the *Coleman* edict. As a reminder, the court in *Coleman* ruled, 'The principle of equal treatment ... applies not to a particular category of person but by reference to the grounds mentioned in Art.1.'[104] The potential is that either statement is so open that it could entertain claims quite the reverse of any purpose that could be ascribed to anti-discrimination legislation. It encompasses persons treated less favourably because of the defendant's *anti*-discrimination conduct. This negative potential was exhibited in two English Court of Appeal cases.

In *Wheeler v Leicester City Council*,[105] a local authority council issued a 12-month bar on a rugby club using the council's recreation ground. This was because three of the club's players participated in a rebel tour of apartheid South Africa, something to which the council objected. The club argued *inter alia* that as the council's conduct was based on 'racial considerations' (opposing apartheid), the *council* was liable for direct racial discrimination. In other words, under Browne-Wilkinson J's wide statement, it no longer mattered if the defendant's conduct was pro- or anti-discriminatory. It was no surprise that the argument was rejected by the Court of Appeal, but only after a renouncement by the statement's author. Browne-Wilkinson LJ, now

101 *English v Sanderson Blinds* [2009] ICR 543 (CA). See above, text to n 86, and further Connolly (n 1 above) text to n 26.

102 *Showboat Entertainment Centre v Owens* [1984] ICR 65 (EAT), discussed further in relation to *Lee v Ashers* in Connolly (n 1 above) text to nn 19 and 65.

103 *Showboat Entertainment Centre v Owens* [1984] ICR 65 (EAT) 73.

104 Case C-303/06 *Coleman*, para 38 (and 50). The 'grounds' alluded to were sexual orientation, religion or belief, disability, and age ('Framework' Directive 2000/78/EC, art 1).

105 [1985] AC 1054 (CA and HL).

sitting in the Court of Appeal, conceded that this literal interpretation of his statement was 'too wide'.[106] Ackner LJ held it could not be used to 'produce consequences totally repugnant to the very purpose of the legislation'.[107]

This did not deter a more radical argument, launched with some chutzpah, in *Redfearn v Serco*.[108] Here, a bus driver, whose passengers numbered mainly Asian, relied on Browne-Wilkinson J's *dictum* to claim that his dismissal for membership of a racist political party was in fact discriminatory. After all, this was 'discriminatory treatment caused by racial considerations'. In other words, a racist was calling in aid of anti-racist legislation *because* he was a racist. The EAT found the logic inescapable, and held that the bus driver was a victim of direct (racial) discrimination.[109] The Court of Appeal reversed, holding that Browne-Wilkinson J's *dictum*, 'does not apply so as to make the employer ... who is pursuing a policy of anti-race discrimination, liable for race discrimination'.[110] In both *Wheeler* and *Redfearn*, the Court of Appeal could not formulate a principle to distinguish such claims, rejecting them only on policy grounds.[111]

Had the *Coleman* and *CHEZ* judgments *and* Opinions considered this English case-law narrative, the court might have been more guarded when promoting the *per se* edict. While it is likely that the ECJ would reject any such dissonant claims under its teleological approach to interpretation,[112] until someone goes to the trouble, time, and expense of arguing a case all the way to Luxembourg, doubts will linger, as the English narrative demonstrated, notably with Mr Redfearn's victory in the EAT.

This is not to say that a 'discrimination *per se*' approach is wholly unwelcome. It has potential to simplify the law and broaden its reach to

---

106 Ibid 1061.

107 Ibid 1060. The matter was not discussed in the House of Lords, who found for the club on an alternative claim that the council had acted *ultra vires*.

108 *Redfearn v Serco (t/a West Yorkshire Transport Service)* [2006] EWCA 659.

109 [2005] IRLR 744 (EAT) especially [30]–[42].

110 [2006] EWCA 659 [43] (Mummery LJ). This goes too far the other way. It does not account for positive action which is unlawful direct discrimination unless sanctioned by the relatively narrow boundaries set by the EA 2010, eg ss 158 and 159. A more generously worded formula regarding Protestant and Roman Catholic employment in Northern Ireland is provided by FETO(NI), art 4. Note that Redfearn won a human rights claim based on political opinion under European Convention on Human Rights, art 11 (Freedom of Association): *Redfearn v United Kingdom* [2013] IRLR 51 (ECtHR).

111 It could have been resolved under the statutory construction rule under *Re Sigsworth* [1935] Ch 89 (Ch) 92 (Clauson J), as something 'obnoxious to the principle'.

112 See eg *R v Henn and Darby* [1981] 1 AC 850 (HL).

encompass unforeseen meritorious claims. But it is a radical departure from orthodoxy which does not appear to have been thought through. Of course, it would prove challenging to draft the edict with precise boundaries, not least because cases such as *Redfearn* and *Wheeler* need to be distinguished from unsanctioned positive action[113] and other benignly motivated discrimination[114] which, policy dictates, *should* attract liability. Thus, the court ought to have provided some policy guidance as to where its boundaries lie, if only to signal that 'anti-purpose' cases will not succeed under its *per se* edict.

While in pockets of Europe, the likes of Mr Redfearn may be encouraged by the *Coleman* edict, back in the UK, a new raft of far-reaching 'associative' claims may emerge, arguing that the *Lee v Ashers*' close-enough rubric no longer applies, or distinguishing it as just one evidential path to a '*per se*' discrimination claim.

'Extended' indirect discrimination

Although the court's deployment of the *Coleman* edict encroached upon the notion of indirect discrimination, it did not entirely erase it from the court's jurisprudence. The judgment confirmed this by producing an alternative finding of indirect discrimination (advising that the practice seemed not to be objectively justified[115]). This was based on the hypothesis that the reason for the 'ostensibly neutral' treatment was not 'based on' ethnicity.[116] Beyond this, no analysis was given, nor theory advanced, in support of this finding (of the practice causing the particular disadvantage). There was no reference to AG Kokott's associative theory.

The particular novel finding was a break with the Directive's conventional definition by including as victims persons not of the relevant ethnic origin. The Directive provides,

> [I]ndirect discrimination shall be taken to occur where an apparently neutral provision, criterion or practice would put persons of a racial or ethnic origin at a particular disadvantage *compared with* other persons ...[117]

Even though the court stated the conventional view that this meant 'it is particularly persons of a given ethnic origin who are at a disadvantage

113 See eg Case C-407/98 *Abrahamsson and Anderson v Fogelqvist* [2000] IRLR 732 (mandatory preference for female candidates).

114 See eg *James v Eastleigh* BC [1990] 2 AC 751 (HL) (free swimming for pensioners discriminated against men, who retired later); *R (A) v Governing Body of JFS* [2009] UKSC 15 (racial preference based on religious doctrine); *Amnesty International v Ahmed* [2009] IRLR 884 (EAT) (protection from violence).

115 Case C-83/14 CHEZ, para 127.

116 Ibid paras 50, 106 and 92–96.

117 Race Directive 2000/43/EC, art 2(2)(b). Emphasis supplied.

because of the measure at issue',[118] its *decision*, classing Roma and non-Roma as one, holds that both groups had suffered a 'particular disadvantage'. Further, there was no indication that both Roma and non-Roma suffered the *same* disadvantage.[119] Indeed, it is reasonable to assume that the offence and stigma suffered by the non-Roma was different in character from that suffered by the Roma, given their history of persecution combined with the stereotyping behind this policy. Such a question was not discussed in the judgment. A clarification would have been welcome. AG Kokott's unacknowledged extended associative theory (illustrated with her nursery scenario[120]) provided a slightly more coherent solution. Her associative theory piggybacked the legislative formula seemingly leaving it intact. However, in substance, either approach challenges the conventional definition of indirect discrimination by including non-members of a suspect class as primary victims with the consequential right to sue. In addition, the *decision* leaves open the possibility that the non-members need not have suffered the *same* disadvantage. The UK legislation more cogently holds to the conventional view on same-group *and* same-harm liability,[121] presenting a potential conflict with Retained EU Law.[122]

As with its approach to direct discrimination, the ramifications of the judgment could be surprising and unwieldy. For instance, in *Hussein v Saints Complete House Furnishers*,[123] a retail furniture store refused to hire youths from the city centre because in the past they attracted unemployed friends who loitered in front of the shop. Compared to other districts in the city, the centre was disproportionately populated with black and Asian residents, one of whom won a claim of indirect discrimination after the store refused to consider him for

118 Case C-83/14 *CHEZ*, para 100.

119 AG Kokott observed that Roma and non-Roma were affected in the same way (Case C-83/14 *CHEZ*, AG98), but drew no rule or boundary from this.

120 See above, text to n 40.

121 Required under EA 2010, s 19(2)(b) and (c) (stating that the claimant must have been put at *that* particular disadvantage suffered by the group); FETO(NI), art 3(2A)(b)(i) and (ii); Sex Discrimination (Northern Ireland) Order 1976, SR 1976/1042, art 3A2(2)(b) and (c); Race Relations (Northern Ireland) Order 1997 SR 1997/869, art 3(1A)(a) and (b).

122 EU employment rights existing before 1 January 2021 were converted into domestic law: EU (Withdrawal) Act 2018, s 7. Under the Trade and Cooperation Agreement 2020, Northern Ireland is bound to follow subsequent EU employment law, while any divergence in the rest of the UK can be redressed with 'rebalancing measures'. See the Trade and Cooperation Agreement, a Summary (UK Government); Explanatory Brochure (EU Commission). More generally, see Malone (n 11 above) parts 10 and 12 respectively.

123 [1979] IRLR 337 (IT).

employment.[124] Upon the *CHEZ* judgment, any white person from that district could sue for indirect discrimination. (If the embargo were motivated by race, under *CHEZ*, the discrimination against the white resident would be *direct*.) The harm could range from being rejected, to being deterred from applying, or just the resultant stigma as a resident. One could conjure up all manner of similarly far-reaching scenarios.[125] In AG Kokott's nursery scenario, for example, in addition to the children deprived of nursery places, the *male* part-time parents, having suffered the same harm as their female colleagues, could sue, even if the females chose not to.[126] This judgment challenges the same-group and same-harm liability principles set out in both the EU and UK legislation. Being insufficiently reasoned and bounded, it could produce unintended consequences.

Collateral damage and standing

The decision favouring the non-Roma claimant may be welcomed by many, but it will sit uncomfortably with those wanting a somewhat tidier underpinning. The fundamental problem common to both the Opinion and judgment is that they are trying to absorb the 'collateral damage' question into a model of substantive law, when it ought to be dealt with as a procedural matter of standing to sue. Instead, it would have been logical (and tidier) to engage the procedural provisions of the Directive, which state that enforcement must be 'available to all persons who consider themselves wronged by failure to apply the principle of equal treatment to them'.[127] This approach is all the more curious because when faced with cases with only hypothetical victims, the court has readily approached the matter via the procedural provisions. Thus, where businesses announced they would never employ immigrants, or homosexuals, the court, after finding that the conduct could amount to direct discrimination (potential applicants

124 See also in the US, where employers advertise in, say, a predominantly white district: eg *USA v City of Warren*, Michigan 138 F 3d 1083 (6th Cir 1998).

125 For this variation, and others, on *Hussein*, see Malone (n 11 above) 152 and (equal pay) 158–161.

126 If the employer had responded to a mother's sex discrimination claim by providing nursery places for the part-time women, the part-time men, now the only ones deprived of nursery care, could 'level up' with a straightforward claim of *direct* sex discrimination. See eg the equal pay case *Hartlepool BC v Llewellyn* [2009] ICR 1426 (EAT) where, following a successful equal pay claim by women in a predominantly female occupation, the men in that occupation were entitled to the same pay as the women.

127 Race Directive 2000/43/EC, art 7(1), located under 'Chapter II Remedies and Enforcement'. For the UK, see nn 132 and 133 below.

could be deterred), considered the standing of an (unharmed) public interest body to bring an action.[128]

The procedural route to providing remedies to those harmed by 'collateral damage' is well established in the US, where the federal equality legislation similarly provides a 'person aggrieved' with the right to sue.[129] For example, two residents (one white, one black) were given standing by the Supreme Court to sue their landlord for its (anti-black) racist policies, causing them a loss of the social and professional benefits of living in an integrated community, as well as the stigma of living in a 'white ghetto'.[130] It is notable here that these plaintiffs did not suffer the same harm as the primary victims and, it is conceivable, not as each other.

Given the far-reaching ramifications, the EU (and UK) courts would be wise to observe this US practice and its limitations.[131] It can resolve other doubts about the *CHEZ* case. First of all, neither the Opinion nor the judgment make clear whether the victim-claimant has to suffer the *same* harm as the principal victims. Short of class actions, this should not be an issue when considering standing. Second, as noted above, *CHEZ* challenges the orthodox approach to direct discrimination as well as the more detailed legislative formula for indirect discrimination. Treating collateral damage claims as a matter of standing may well provide British courts with a means to avoid the complexities of compatibility and interpretive issues post-Brexit. The EA 2010's enforcement provisions are expressed quite passively, providing jurisdiction to courts and tribunals to determine complaints

128 Respectively, Case C-54/07 *Feryn* [2008] ECR I-05187 (ECJ); C-81/12 *Asociatia ACCEPT v Consiliul National pentru Combaterea Discriminarii* [2013] ICR 938 and C-507/18 *NH v Associazione Avvocatura per I diritti LGBTI – Rete Lenford* [2020] 3 CMLR 33.

129 See eg Fair Housing Act 1968 (formally Civil Rights Act 1968), s 810(a) (codified as 42 USC s 3610(a)); Civil Rights Act 1964, Title VII (employment), s 706 (42 USC s 2000e-(5)(f)(1)), although courts often refer to 'associative theory' at the same time: see eg *Clayton v White Hall School* 778 F 2d 457 (8th Cir 1985) 459.

130 *Trafficante v Metropolitan Life Insurance Co* 409 US 205 (1972, US Sup Ct) 208. The notion has been applied to employment. In *Angelino v New York Times* 200 F 3d 73 (3rd Cir 2000), the employer stopped hiring when it reached the first woman's name on its priority list; as a well as the women, the *men* below that name not hired had standing to sue for the sex discrimination against the women.

131 See eg *Thompson v North American Stainless* 131 S Ct 863 (2011, US Sup Ct) 868–870, limiting the reach of the 'person aggrieved' requirement; *Lyman v Nabil's* 903 F Supp 1443 (D Kan 1995) 1446, distinguishing *Trafficante* as not extending to sex discrimination; *Patee v Pacific Northwest Bell Tel Co* 803 F 2d 476 (9th Cir, 1986) 479 (men did not have standing to protest their depressed wages allegedly resulting from their employer's discrimination against women in the same job classification).

'relating to a contravention' of parts of the Act, covering employment, services and public functions, premises, education, associations, or other ancillary matters.[132] This is not so for Northern Ireland, where the enforcement provisions state that the complainant is the victim of the alleged discrimination.[133] Here then, 'collateral damage' victims must rely on the logic of *CHEZ*.

The missteps in *CHEZ*

One can assume that both the Advocate General and the court were driven by the policy concern of the wholesale treatment rooted in racial stereotyping, and that the only remedy was via a sole complainant who happened to be non-Roma. However, this result was achieved via many missteps in a patchwork of incomplete notions and unconnected strands. It is unsurprising then, that the case may produce some unintended consequences. The principal missteps were:

1. *'Extended' associative discrimination*. Advocate General Kokott's Opinion deployed models of discrimination as generally understood in EU and UK jurisprudence. The misstep was attaching (or 'piggybacking') her associative theory, which had more to do with associated harm than anything else. Applied to indirect discrimination, it had extraordinary potential. In this case, embracing non-members of a suspect class as primary victims, a matter for standing, not substantive law.
2. *Unorthodox comparison*. The court's finding of direct discrimination was facilitated by a comparison that could do no more than distinguish victims from non-victims, irrespective of ethnicity.
3. *A suggestion of an intentional/non-intentional model*. The court's apparent reliance on some unspecified level of motive departed from its conventional form-based model and encroached upon the conventional model of indirect discrimination set out in the legislation. Neither consequence was acknowledged in the judgment. If followed, this aligns EU law with the US 'intentional/non-intentional' model, but with no complementary framework regarding such matters as benign motives, the pretext doctrine, its shifting burdens, and the separate provisions on standing.
4. *Adoption of the* Coleman *edict*. The edict, a blank canvas, was speckled with varying notions including 'collateral victims', an orthodox comparison, and conduct ranging from that 'motivated by' ethnicity to that merely 'relating to' ethnicity. The edict has

132 EA 2010, s 120 (employment tribunals) and s 114, respectively.

133 See eg FETO(NI), art 38; Race Relations (Northern Ireland) Order 1997, SR 1997/869, art 52.

extraordinary potential, notably for *Sanderson*-type cases. It was adopted without boundaries, leaving its potential unspecified and vague.

5 *Extending the reach of indirect discrimination*. This again was done in the face of the contrary legislative formula, with no reasons given. There was no acknowledgment of this, nor the resulting extraordinary potential.

6 *Absorbing procedural matters into substantive law*. In doing this, both the Advocate General and the court have distorted the established models of discrimination and have sown doubt into their exact meaning.

The associative myth of *CHEZ*

The case has been heralded as an example of associative discrimination.[134] This is an error. Even where this term was used, it was unsustainable. AG Kokott's theory for direct discrimination did no more than restate the simple examples, which are no more than treatment 'on grounds of' race (or other protected characteristic). Her 'restaurant' example was her most ambitious illustration, but goes no further than the *Lee v Ashers*' restrictive close-enough rubric. Labelling such cases as associative perpetuates the myth and risks narrowing the scope of the direct discrimination by restricting the reach of the 'grounds of' element.

By contrast, the Advocate General's application of her test for indirect discrimination was extraordinarily wide, equating her associative theory with 'collateral damage',[135] and thus encompassing anyone harmed by the discriminatory act. This permits, say, men affected by less favourable treatment of (predominantly female) part-time workers to claim that they have been discriminated against 'by association'. The myth of association is more readily understood with large-scale examples. Suppose a national rule disadvantaging part-time workers by requiring longer service to acquire unfair dismissal and redundancy rights.[136] This would obviously put women at a particular disadvantage. Under AG Kokott's associative theory, any affected man in the country could sue for 'associative' sex discrimination. It is somewhat creative to hold here even a non-personal 'link' or association of any meaning save for the shared or 'collateral' damage. In fact, the only explanation for liability under this theory is collateral damage, which is not a form

134 See n 11 above.

135 Case C-83/14 *CHEZ*, para AG58, applied to indirect discrimination AG106–AG109.

136 *R v Secretary of State for Employment, ex p EOC* [1995] 1 AC 1 (HL).

of discrimination. As such, the reason for liability must be found elsewhere, with the likely route being procedural.

Meanwhile, the court's findings of direct and/or indirect discrimination required no supplementary theories piggybacking the conventional models of discrimination. Instead, the judgment ripped up the conventional form-based models. In doing so, it eradicated any notion that associative discrimination was a term of art in EU jurisprudence. Again, anyone harmed by the discriminatory conduct was included. No 'association' beyond this was required.

CONCLUSION

Associative discrimination has two legislative origins in modern discrimination law. One, from the US, is a narrow but manageable reasoning ultimately dependent upon the targeted victim-claimant's protected characteristic. Its limited reach is down to the 'closed' (US federal) legislative formula. The other, stemming from the more open EU (and UK) formulas, has no such limitation. This was recognised in the 1970s by the English Court of Appeal, and subsequently, the House of Lords, both observing that associative scenarios fell within the language of the UK open formula, as well as its mischief, or purpose.[137] Its open formula meant that the claimant's protected characteristic need not have been a factor. In time, the English judiciary learnt that it had to impose some boundaries on this open formula, save it facilitate 'anti-purpose' cases.

The reasoning offered in *CHEZ* (and subsequently in *Lee v Ashers*) appeared to have no knowledge of either origin. The Advocate General produced an extended theory of associative discrimination, which was wrongly equated with 'collateral damage', confused with standing, and in any case vanished under the court's overbroad models of discrimination. The *Coleman* edict, on which it seemingly relied, is a single incomplete and unbounded model of direct discrimination. Without qualification, it can supplant the form-based conventional and established formulas, import procedural matters into substantive law, and even entertain 'anti-purpose' claims or benign motive defences. Without qualification, notably, a conventional comparison, it will produce more problems than it apparently solved. The case leaves open the *possibility*, but no more, that the *Sanderson* principle of extended liability for harassment could apply to direct discrimination.

137 *Applin v Race Relations Board* [1973] QB 815 (CA) 828 and 831 (Stephenson LJ), affirmed [1975] 2 AC 259 (HL) 289–290 (Lord Simon), commenting on the similarly formulated version in the Race Relations Act 1968, s 1(1). See further Connolly (n 1 above), text to nn 1, 2 and 37.

The advice on indirect discrimination suggests that third parties, not belonging to a suspect class but suffering under the same facially neutral treatment, can claim that they suffered *discrimination*; this is irrespective of whether any primary victims complain. This presents a particular challenge to the UK statutory provisions, which express that only victims with the relevant protected characteristic can sue for indirect discrimination.

More generally, it is received wisdom in the UK by scholars and judges alike (including the Supreme Court), that *CHEZ* is a case of associative discrimination.[138] It is no such thing. This case – a patchwork of incomplete notions and unconnected strands – was *in fact* a straightforward example of indirect discrimination requiring resolution for the non-Roma claimant under the procedural provisions on standing. For the ECJ, it was a scenario falling within its *Coleman* edict and/or a US-style 'intent' model, *or* a redefined extended model of indirect discrimination. No association was required or mentioned. The one certainty about the radical judgment is that it did not endorse as a term of art the Advocate General's extraordinary associative theory, nor indeed, *any* associative theory. This vanishing act was one of the few positives emerging from the judgment. It is unfortunate that this also erased many boundaries of the established principles of discrimination. As such, truly it was a *great* vanishing act.

What is required now is for the ECJ to reinstate the orthodox models of direct and indirect discrimination and impose some boundaries on the Coleman edict. If EU law is to adopt fully a US-style model and its complementary framework, it should not be done without extensive consideration. Whatever the outcome, any reform should continue to distinguish between substantive law and standing, announce the necessary policy boundaries, and, of course, not engage with notions of associative discrimination as terms of art.

138 See n 11 above.

Northern Ireland
Legal Quarterly
Autumn Vol. 72 No. 3 (2021) 542–568
Article DOI: 10.53386/nilq.v72i3.980

Formulating the legislative structure of a hate crime*

Jennifer Schweppe

University of Limerick
Correspondence email: jennifer.schweppe@ul.ie

ABSTRACT

While hate crime legislation is well established in England and Wales, Scotland, and Northern Ireland, Ireland has failed to address the issue of hate crime on a statutory basis. Law reform processes are currently underway across these jurisdictions, and this article seeks to explore a fundamental question in this context, that is, the relative merits of various approaches to structuring hate crime legislation.

Keywords: hate crime; hate studies; sentencing; law reform.

INTRODUCTION

Across the four jurisdictions on the two islands – Ireland, Northern Ireland, Scotland, and England and Wales – law reform efforts are underway to determine the means by which the hate element of a crime should be addressed by the law. Interestingly, all four jurisdictions currently take very different approaches to the issue, from Ireland, which relies purely on judicial discretion, to England and Wales, which has what might be regarded as the most sophisticated approach to hate crime globally. Scholars have identified that there are three core questions whose relevance transcends the differences and commonalities in the construction of hate crime laws:

- the range of protected categories included in such legislation;
- the formal recognition of the hate element as either part of an aggravated offence or through the sentencing process;
- the extent to which the hate element should be present in the offence.[1]

Thus, legislating against hate is not a simple task: Rosenberg observes that, for better or worse, 'certain bias crimes represent a drastic

* First published in NILQ 72.AD2 (2021) ADVANCE 2 August 1–27.

1 See, for example, Kay Goodall, 'Conceptualising "racism" in criminal law' (2013) 33(2) Legal Studies 215; Jon Garland and Neil Chakraborti, 'Divided by a common concept? Assessing the implications of different conceptualizations of hate crime in the European Union' (2012) 9(1) European Journal of Criminology 38.

doctrinal departure from a longstanding maxim of criminal law'.[2] The reason for this is that, traditionally, the motivation of the offender is dealt with at sentencing, rather than being included in a substantive offence – thus addressing the third of the commonalities above. For this reason, developing the very structure of hate crime legislation is perhaps one of the more legally contentious questions we must ask: do we keep with tradition and consider the hate element at the sentencing stage; do we insert it into the offence itself; or do we create a 'third way' of ensuring the hate element is presented in the case? The first approach – the 'enhanced sentencing' approach – provides through statute that, where a hate element is present in a case, the court must or should treat that element as an aggravating factor in sentencing. The second approach – the 'aggravated offences approach – creates new (aggravated) forms of existing offences, by attaching the hate element to the base offence as well as (typically) providing a higher maximum penalty for the aggravated offence than for the base offence. The third approach – referred to in some of the literature as the 'hybrid approach' – is to create a separate charge for the hate element of the offence which can be attached to any offence.[3] The fourth and final model is the penalty enhancement statute, common in codified systems of law, which typically treats the hate element as an aggravating factor in sentencing, whilst simultaneously enhancing the penalty which can be imposed for all offences, by increasing the maximum sentence which can be imposed, or setting up a specific enhancement to be attached to the sentence.

While there are many theoretical debates to be had regarding the necessity or justification for hate crime laws, a practical consideration which must be considered is whether the legislation will 'work': that is, will the legislation be used to ensure that hate crime is appropriately investigated, prosecuted and sentenced, or will the hate element of the crime remain 'disappeared'. This article, then, will explore the current approaches to hate crime of all four jurisdictions on these two islands, seeking to understand how such approaches might guide and inform law reform processes across the two islands. It will look to four key approaches: relying on judicial discretion; introducing statutory sentence enhancement provisions; the Scottish approach of having

2 Michael T Rosenberg, 'The continued relevance of the irrelevance-of-motive maxim' (2008) 57 Duke Law Journal 1143, 1173.

3 The term 'hybrid approach' is one coined by Goodall and Walters (Kay Goodall and Mark Walters, *Legislating to Address Hate Crimes against the LGBT Community in the Commonwealth* (Human Dignity Trust 2019)) and adopted by Desmond Marrinan (Independent Review of Hate Crime, *Hate Crime Legislation in Northern Ireland – An Independent Review Consultation Paper* (Hate Crime Legislation in Northern Ireland 2020), and for the sake of consistency within the literature on this issue, I also use it here.

a sentencing-related charge; and the introduction of new aggravated offences. In exploring the benefits of each model, I look to the clarity of the law, its efficacy and the impact on offenders. Though the benefits of each model are clear, I ultimately advocate for the introduction of the aggravated offences model.

There is also a European context to this issue. The European Union (EU) Council Framework Decision on combating certain forms and expressions of racism and xenophobia by means of criminal law (the Framework Decision) was introduced in 2008. The stated purpose of the Framework Decision is to ensure that 'certain serious manifestations of racism and xenophobia are punishable by effective, proportionate and dissuasive criminal penalties throughout the EU', and further aims to 'improve and encourage judicial cooperation' in this context. Article 4 of the Framework Decision addresses the issue of hate crime, and in this regard requires member states to 'take the necessary measures to ensure that racist and xenophobic motivation is considered an aggravating circumstance or alternatively that such motivation may be taken into consideration by the courts in the determination of the penalties'. The report on the implementation of the Framework Decision elaborates, stating that the member states must ensure 'that racist and xenophobic motives are properly unmasked and adequately addressed'.[4]

The EU Fundamental Rights Agency observes that this requirement under article 4 reflects the rights of victims of racist crime as established and required by case law of the European Court of Human Rights (ECtHR). The Court in *Angelova and Iliev v Bulgaria*[5] set out the obligations of the state in relation to a racially motivated murder of two members of the Roma community. In the context of an application under article 2 in conjunction with article 14, the court set out the obligations of states:

> … when investigating violent incidents State authorities have the additional duty to take all reasonable steps to unmask any racist motive and to establish whether or not ethnic hatred or prejudice may have played a role in the events. Failing to do so and treating racially induced violence and brutality on an equal footing with cases that have no racist overtones would be to turn a blind eye to the specific nature of acts that are particularly destructive of fundamental rights.[6]

---

4 European Union Agency for Fundamental Rights, Opinion of the European Union Agency for Fundamental Rights on the Framework Decision on Racism and Xenophobia – with special attention to the rights of victims of crime (European Union Agency for Fundamental Rights 2013) 8.

5 App no 55523/00 (ECtHR, 26 July 2007).

6 Ibid para 115.

Importantly, as Hanek observes, while having specific legislation which addresses the hate element of a crime is beneficial, the obligation to unmask the hate element applies even in the absence of such legislation.[7] It is also important to note that the efficacy of hate crime legislation is not entirely dependent on its structure or scope: research shows, and advice from the European Commission tells us, that in the absence of structural and policy supports to bolster the legislation being implemented at a national level, the hate element of a crime disappears though the criminal process.[8] Legislation is just one element to facilitate the process, but that element cannot be underestimated: legislation will inform and drive these supportive and facilitative processes. The first part of this article will explore the manner in which courts understand and approach hate crime in the absence of legislation using Ireland as a case study. The next will explore the three legislative approaches that can be considered: aggravated sentencing provisions; aggravated offences; and the so-called hybrid model.

MODEL 0: IRELAND – UNDERSTANDING HATE IN THE ABSENCE OF LEGISLATION

While most European countries have introduced hate crime legislation, Ireland remains an outlier, with no statutory recognition of hate crime and limited jurisprudence on the issue. The latter fact is unsurprising, given the fact that the sentencing process in Ireland is a discretionary one, with few limitations and even less guidance given either by the legislature or the appellate courts on sentencing issues: as O'Malley states, 'Ireland's sentencing system remains largely discretionary, reflecting a commitment to individualised justice for criminal offenders.'[9] It was only very recently that the Irish courts addressed the sentencing of racist crime in any way. The first written judgment in which the question as to whether a racist motivation is an aggravating factor was given in *Director of Public Prosecutions v Elders*.[10] In the case, the racist element was present at the beginning of a series of events which took place where the appellant said to the injured party: '"eff off"... "eff off Packi [sic] bastards"'. The sentencing judge assessed the offence as being at the top end of seriousness, and that 'the racist

7 Aleš Gião Hanek, 'International legal framework for "hate crimes: which law for the "new" countries' in Amanda Haynes, Jennifer Schweppe and Seamus Taylor (eds), *Critical Perspectives on Hate Crime: Contributions from the Island of Ireland* (Palgrave Macmillan 2017) 467.

8 Jennifer Schweppe, Amanda Haynes and Mark A Walters, *Lifecycle of a Hate Crime: Comparative Report* (Irish Council for Civil Liberties 2018).

9 Thomas O'Malley, *Sentencing Law and Practice* 3rd edn (Round Hall 2016) 1.

10 [2014] IECA 6.

element was an aggravating factor' and sentenced the appellant to a term of five years' imprisonment, the maximum sentence available for that offence.

In assessing whether the sentence imposed was appropriate, Birmingham J discussed the aggravating factors:

> Among the very many aggravating factors present were that there was a racist dimension, an aspect that was very properly highlighted by the Circuit Court judge. It may be that as counsel for the appellant said that this was not the case where someone was attacked because of their race, but that there was a racist dimension is nonetheless clear and that is an aggravated fact.

While accepting the very serious nature of the offence, the Court of Appeal found that the sentencing court had failed to take appropriate account of the mitigating factors and suspended the final 12 months of the sentence, subject to an offer of €4000 compensation being paid to the injured party.

Whilst there have been some – though no more than a handful – of reported cases since *Elders* which considered racism as an aggravating factor, *Elders* is considered the core precedent on the issue. However, it leaves a number of questions unanswered: as discussed above, two of the key issues which hate crime legislation addresses are (1) the personal or protected characteristics relevant in the context of such legislation; and (2) the extent to which the hate element must be present in the offence (eg whether motivation of hostility or demonstration of hostility is required) . *Elders* offers no advice on either of these issues, even by way of *obiter* statements. With respect to the first question, in the context of disablist hate crime, Kilcommins et al[11] observe that, while there is little jurisprudence on the question, there is 'no reason why a sentencing judge in Ireland could not regard the fact that the crime was committed against a person with a disability as an aggravating factor'.[12] That said, there is, of course, nothing *requiring* a court to take it into account as an aggravating factor either. In this context, Kilcommins et al recommend that a statutory provision be introduced which 'provides that an offence committed against a vulnerable person such as a person with a disability may be considered an aggravating factor at sentencing stage'.[13] The same line of argument applies in the context of other commonly protected characteristics, such as sexual orientation, gender identity or expression, religion, or age.

---

11 Shane Kilcommins, Claire Edwards and Tina O'Sullivan, *An International Review of Legal Provisions and Supports for People with Disabilities as Victims of Crime* (Irish Council for Civil Liberties 2013).

12 *Elders* (n 10 above) 51

13 Ibid 228.

With respect to the second question, it is not clear what level of proof is required in order to establish the racist element, and this issue has not been clarified by later reported cases. In *Director of Public Prosecutions v Collins*,[14] for example, the trial judge seems to have taken into account the fact that the offence 'may have been racially motivated'.[15] Birmingham J stated:

> He was prompted to do this by a sentence in the probation report which quotes their client as saying 'he (that is the accused) says he watched two foreign nationals cross the road to his girlfriend.' By reference to this sentence the judge said that he felt that it was highly probable that the attack had some element of racism to an unspecified degree.[16]

The Court of Appeal did not take the opportunity to consider whether this amounted to proof of racist motivation on the part of the accused, nor whether such evidence was appropriate to consider as proof of a racist motivation. While the court did not explicitly criticise the sentencing judge for treating statements in the probation report as proof of a racist motivation to the offence, it did state that it was 'not clear' what role, if any, this concern regarding a racist motivation had when it came to determining the sentence.

Aside from the legal issues which arise, there are also practical issues relating to the absence of hate crime legislation in an Irish context. Quite simply, the Irish criminal justice process has been shown to be incapable of addressing or even recognising hate crime in the absence of legislation. Indeed, Haynes and Schweppe have clearly shown that the hate element of a crime is 'disappeared' from the process as the offence makes its way through the criminal justice system.[17] So, while the understanding of the courts of hate and hate crime is not terribly sophisticated, this lack of understanding is matched across the process, and, indeed, it is only in rare cases that the court will have an opportunity to review the hate element of a crime.[18]

I believe that the Irish situation is currently untenable. In the absence of clear guidance from the appellate courts, offenders are labelled as criminal racists where the offence 'may have been racially motivated', a standard of proof far too low. The lack of clarity regarding the range of categories to which the aggravation applies is equally problematic. After years of inaction and outright rejection of the claim that hate

14 [2016] IECA 35.
15 Ibid para 15.
16 Ibid.
17 Schweppe et al (n 8 above).
18 Ibid.

crime was a problem in Ireland, the state has come to accept that the introduction of hate crime legislation is required.[19]

Indeed, the question as to whether introducing hate crime legislation in this context would be useful or not is one which has not just troubled Ireland, but is a well-rehearsed issue globally.[20] It is now generally accepted that, in order to address hate crime, legislation is required to ensure that the criminal justice process responds to the phenomenon effectively.[21] One of the key questions which then remains is what structure such legislation should take. The remainder of this article will reflect upon the manner in which three key legislative provisions operate in practice and consider each from a law reform perspective. The aim is to inform the law reform processes across all four jurisdictions, though of course cultural, legal and policy differences will influence the ultimate recommendations for legal developments in each jurisdiction. Across Northern Ireland, England and Wales, and Scotland, three models of legislation are in operation. These three models will now be considered in turn.

MODEL 1: ENGLAND AND WALES – AGGRAVATED OFFENCES

While legislators and courts in the United States have been grappling with concepts and constructions of hate crime for decades, they have a more recent pedigree in England and Wales.[22] The initial legislative vehicle for recognising hate crime was part 2 of the Crime and Disorder Act 1998. It introduced the concept of the racially aggravated offence which carried a higher penalty than its non-racially motivated counterpart.[23] As Malik notes, the Act does more than simply bolt on the aggravating factor to the existing offences: rather, 'the new racially aggravated offences are aimed at conduct which causes harm of a qualitatively different type to that caused by the basic offences'.[24]

19 See Department of Justice, *Legislating for Hate Speech and Hate Crime in Ireland Report* (Department of Justice 2020).

20 It is not proposed to explore these arguments in any detail in this article given its focus, but on this issue see, for example, Benjamin Bowling and Coretta Phillips, *Race, Crime and Justice* (Pearson Education 2002); James B Jacobs and Kimberly Potter, *Hate Crimes* (Oxford University Press 1998).

21 See, for example, Schweppe et al (n 8 above).

22 Neil Chakraborti and Jon Garland, *Hate Crime: Impact, Causes and Responses* (Sage 2015).

23 In England and Wales, a model of aggravated sentencing is also used: this section will explore the aggravated offences only for the purposes of illustrating the legislative approaches.

24 Maleiha Malik, 'Racist crime: racially aggravated offences in the Crime and Disorder Act 1998 part II' (1999) 62 Modern Law Review 409, 419.

Section no	Offence	Max penalty non-aggravated	Max penalty aggravated
OAPA, s 20	Malicious wounding/grievous bodily harm	5 years	7 years
OAPA, s 47	Actual bodily harm	5 years	7 years
CJA, s 39	Common assault	6 months	2 years
CDG, s 1	Criminal damage	10 years	14 years
POA, s 4	Fear of provocation of violence	6 months	2 years
POA, s 4A	Intentional harassment, alarm or distress	6 months	2 years
POA, s 5	Harassment, alarm or distress	£1000 fine	£2500 fine
PHA, s 2	Harassment	6 months	2 years
PHA, s 2A	Stalking	6 months	2 years
PHA, s 4	Putting people in fear of violence	5 years	7 years
PHA, s 4A	Stalking involving fear of violence or serious alarm or distress	10 years	14 years

Table 1: Offence type and maximum penalty (hate and non-hate aggravation) under Crime and Disorder Act 1998.[25] **Key:** OAPA = Offences Against the Person Act 1861; CDG = Criminal Damage Act 1971; PHA = Protection from Harassment Act 1997; CJA = Criminal Justice Act 1988; POA = Public Order Act 1986

The Act, under sections 28–32, created new forms of racially aggravated offences, which include assault, assault occasioning actual bodily harm, malicious wounding/grievous bodily harm, harassment and stalking, as well as various public order offences (see Table 1). The Act was later amended to include religiously aggravated offences under the Anti-terrorism, Crime and Security Act 2001. The Act does not use the term 'hate crime', but rather addresses the hate element by reference to the identity characteristics protected, for example in the creation in section 29 of 'racially or religiously aggravated assault'. Section 28 of the Crime and Disorder Act states:

> (1) An offence is racially or religiously aggravated for the purposes of sections 29 to 32 below if—
>
> (a) at the time of committing the offence, or immediately before or after doing so, the offender demonstrates towards the victim of the offence hostility based on the victim's membership (or presumed membership) of a racial or religious group; or

25 Table adapted from Law Commission, *Hate Crime: The Case for Extending the Existing Offences – A Consultation Paper* (Law Com CP No 213 2014) 24.

> (b) the offence is motivated (wholly or partly) by hostility towards members of a racial or religious group based on their membership of that group.

Following this definition, in the offences set out in sections 29–32, the maximum sentence available to the court in sentencing, for example, racially aggravated assault, is increased as compared to the sentence available on a non-aggravated charge. For some offences, the penalty enhancement is substantial, as is evident in Table 1. For example, the offence of assault currently carries a maximum sentence of six months' imprisonment which is increased by 400 per cent to two years imprisonment for cases aggravated by racial or religious hostility.

It is useful to note in this context that the Criminal Justice Act 2003 in England and Wales utilises a sentence enhancement model to address hate crime in relation to hate crime against individuals relating to their transgender identity, disability, or sexual orientation, thus creating a hierarchy of victims in that jurisdiction. The Law Commission of England and Wales conducted a consultation process and then published a report on hate crime in that jurisdiction in 2013–2014.[26] In the context of the present discussion, the primary question was whether the aggravated offences should be extended to the grounds of hostility protected under the 2003 Act. The Law Commission highlighted that one of the primary advantages of having aggravated offences as compared to the enhanced sentencing model is the fact that the former carry a 'unique descriptor', reflecting the fact that aggravated offences are considered more serious than their basic counterparts:

> The 'aggravated' label is designed to carry and communicate a stigma which 'stings' more deeply than the mere fact of conviction for the basic offence, even with an enhanced sentence.[27]

In this context, in its Consultation Paper, the Commission highlighted the fact that, with aggravated offences, the label will attach to the offender's criminal record. The offences also, as the Commission highlights in its Report, can be seen as giving recognition to 'the particular seriousness of hate crime, the greater culpability of its perpetrators and the greater harms it can cause'.[28] While in the Consultation Paper the Commission seemed in favour of extending the offences, it took a more cautious view in its Report, ultimately suggesting that a wider review of the aggravated offences was necessary, but that in the absence of such a review, the offences should be extended to the crimes committed on

---

26 For details of this process, see Law Commission, 'Hate Crime'.

27 Law Commission (n 25 above) 68.

28 Law Commission, *Hate Crime: Should the Current Offences be Extended?* (Law Com No 348, 2014) 96.

the basis of sexual orientation, disability and transgender identity, primarily for reasons based on ensuring equality across the grounds.

Linked to this issue, in a report published before the Consultation Paper, Burney and Rose note that defence lawyers emphasised the 'vehemence' with which racial aggravation was denied by defendants, highlighted in the fact that, in 1999, not-guilty pleas were entered for over 83 per cent of racially aggravated offences, and for only 47 per cent of substantive offences.[29] In its 2014 Consultation Paper, the Law Commission observed that one of the problems identified in prosecuting hate crime is the fact that defendants will plead only to the non-aggravated form of the offence, leading to charges being downgraded or even dropped.[30] However, for England and Wales, the conviction rate for racially and religiously aggravated offences is relatively high at 83.8 per cent.[31] Yet, in their recent study, Walters et al note that, when a more holistic analysis of the process is conducted, the very high conviction rates as set out by the Crown Prosecution Service are flattened somewhat.[32] They observe that prosecution outcomes for disability hate crime (73.13%) and religious hate crime (79.1%) are lower than homophobic and transphobic (82.98%) and race hate crime (84.07%).[33] The rate of guilty pleas also varies across categories, with only 63.44 per cent guilty pleas for disability hate crime, and 74.27 per cent guilty pleas for racist hate crime.[34]

In its Report, the Commission discussed the arguments in favour of extending the offences under 10 headings, the first of which related to the need to treat the protected characteristics equally.[35] The other nine, however, addressed the function of aggravated offences in the criminal justice process, including: labelling; the communicative and deterrent effects of aggravated offences; the potentiality for increased public confidence; and other procedural aspects. A number of arguments against the extension of the offences were also addressed, including: the current complexity of aggravated offences; the interrelationship between the aggravated offences and aggravated sentencing provisions; and the adequacy of sentencing provisions to address the mischief.

Despite these extensive considerations, Bakalis helpfully observes that there were ultimately two primary reasons the Commission was

29 Elizabeth Burney and Gerry Rose, *Racist Offences: How Is the Law Working? The Implementation of the Legislation on Racially Aggravated Offences in the Crime and Disorder Act 1998* (Home Office 2002) 89–90.

30 Law Commission (n 28 above) 131.

31 Hate Crime Annual Report 2016–2017 (Crown Prosecution Service 2016–2017) 8.

32 Mark Walters, Susann Wiedlitzka and Abenaa Owusu-Bempah with Kay Goodall, *Hate Crime and the Legal Process: Options for Law Reform* (University of Sussex 2017).

33 Ibid 60.

34 Ibid.

35 Law Commission (n 28 above).

swayed in its view between consultation and report stage.[36] The first relates to the benefits of the proposed extension. While it was accepted that there may well be symbolic, communicative and fair-labelling benefits to the extension, it was also accepted that these benefits are speculative, and could potentially be achieved through the development of the enhanced sentencing regime. Second, and Bakalis suggests, most importantly, the consultation process brought to light a number of procedural and practical problems in relation to the operation of the aggravated offences which undermine their effectiveness:

> The unduly complex nature of the offences which allow for either the demonstration of hostility or the motivation of hostility causes problems in practice for prosecutors, and results in plea-bargaining, or the dropping or downgrading of aggravation charges. In many cases, it has also led to the aggravated charges not being brought in the first place as they are deemed too difficult to prosecute.[37]

Bakalis observes that the combined effect of these practical problems has led to aggravated offences not being used 'effectively', and not ultimately achieving the purposes for which they were designed.[38]

In their most recent study, Walters et al utilised rich empirical data to highlight some of the issues relating to the current legislative models in England Wales. In that study, they observe that the aggravated offences provisions were considered to be the 'cornerstone of the legal framework', and they found that those provisions were 'generally well comprehended by most practitioners, including judges'.[39] Indeed, the most significant criticism of the Crime and Disorder Act 1998 was that it did not apply across all protected characteristics.[40] A new issue identified in their research was the perception by some barristers and judges that the Crown Prosecution Service was engaging in 'over-charging', and had adopted an 'overly-zealous "pro-charge" policy'.[41] Though they highlighted a number of procedural issues which can cause 'injustice and unfairness', which their research uncovered with reference to the operation of the aggravated offences provisions of the Crime and Disorder Act, they were of the view that it was possible to rectify these procedural problems. Ultimately, when compared with

36 Chara Bakalis, 'Legislating against hatred: the Law Commission's report on hate crime' (2015) Criminal Law Review 192, 201.

37 Ibid 201–202.

38 Ibid 202.

39 Mark Austin Walters, Abenaa Owusu-Bempah and Susann Wiedlitzka, 'Hate crime and the "justice gap": the case for law reform' (2018) 12 Criminal Law Review 961, 972.

40 Ibid.

41 Abenaa Owusu-Bempah, Mark Walters and Susann Wiedlitzka, 'Racially and religiously aggravated offences: "God's gift to defence"?' (2019) 6 Criminal Law Review 463, 473.

the aggravated sentencing model, they conclude that 'the rights and interests of defendants can be better protected through prosecution of specific hate crime offences (particularly where there is a trial by jury) than through the application of enhanced sentencing provisions.[42]

In its most recent Consultation Paper, *Hate Crime Laws: A Consultation Paper*[43] the Commission highlighted a number of criticisms regarding the legislation. What is interesting to note in this context is that no explicit concerns were expressed with respect to the legislative structure, other than the disquiet regarding the disparity in the way in which groups were treated under the legislation. One justification given by the Commission for retaining aggravated offences was to ensure that the increased maximum sentence available under those provisions was still available, though it admitted that the sentences imposed rarely exceed the sentence for the non-aggravated version. The Commission went on, however to state that adopting enhanced sentencing only 'is likely to send the wrong message ... and undermine the overall deterrent effect of hate crime laws'.[44]

This most recent Law Commission Consultation Paper referred to aggravated offences as 'among the most powerful forms of condemnation of characteristic-based criminal hostility'.[45] It further noted that repealing these provisions and relying only on the aggravated sentencing model would be problematic, 'particularly as one of the key purposes of hate crime laws is to signal the unacceptability of this conduct'.[46] Ultimately, the Commission provisionally proposes in this Consultation Paper that aggravated offences be retained, given their symbolic and deterrent effect. It is unclear in the Paper, however, the extent to which this recommendation is made given the symbolic effect of *repealing* such provisions, or whether it is based on the inherent deterrent and symbolic effect of legislation.

MODEL 2: NORTHERN IRELAND – ENHANCED SENTENCING

While the substance of legislative developments in the area of crime and criminal justice matters in England and Wales is often followed in Northern Ireland a few years later by domestic legislation which takes

42 Ibid 484.
43 Law Commission, *Hate Crime Laws: A Consultation Paper* (Law Com CP 250, 2020).
44 Ibid 175.
45 Ibid 381.
46 Ibid.

account of any local differences applicable in Northern Ireland,[47] it was not until 2004 that legislation was introduced in Northern Ireland to address hate crime. Article 2 of the Criminal Justice (No 2) (Northern Ireland) Order 2004 provides that, where an offence was aggravated by hostility, the court must treat that as an aggravating factor in sentencing, which increases the seriousness of the offence, and must state in open court that this is the case.[48] Here the legislation is based on the model from England and Wales, providing that the offence is aggravated by hostility if:

> ... at the time of committing the offence, or immediately after doing so, the offender demonstrates towards the victim of the offence hostility based on:
>
> (i) The victim's membership[49] (or presumed[50] membership) of a racial group;[51]
>
> (ii) The victim's membership (or presumed membership) of a religious group;[52]

47 The Crime and Disorder Act 1998, for example, was not extended to Northern Ireland 'because of the technical difficulties of doing so which made it impossible either to extend directly the provisions in their entirety or to introduce them by negative resolution procedure'. See *Race Crime and Sectarian Crime Legislation in Northern Ireland: A Consultation Paper* (Northern Ireland Office 2002).

48 For context, it is important to note that as well as considering the question as to whether the range of aggravated offences should be extended, the Law Commission in England and Wales considered the operation of the enhanced sentencing system in that jurisdiction. Legislation in England and Wales does not provide for specific offences where hostility is demonstrated towards a victim's sexual orientation, gender-identity or disability. Instead, s 146 of the Criminal Justice Act 2003 provides for sentencing provisions allowing judges to increase the penalty for an offender where there is evidence that proves he or she demonstrated hostility towards the victim based on the victim's sexual orientation, transgender identity and/or disability. Importantly, there is no corresponding aggravated offence in the context of these protected characteristics. Further, the Criminal Justice Act 2003 provides that where a court is considering the seriousness of an offence other than one provided for in ss 29 to 32 of the Crime and Disorder Act 1998, and the offence is racially or religiously aggravated, the court must treat that as an aggravating factor, and state in open court that the offence was so aggravated.

49 'Membership' includes association with members of that group, art 2(5).

50 'Presumed' means presumed by the offender, art 2(5).

51 'Racial group' means a group of persons defined by reference to colour, race, nationality or ethnic or national origins, and references to a person's racial group refer to any racial group into which he falls. See art 2(5) and art 5, Race Relations (Northern Ireland) Order 1997 (SI 1997/869, NI 6).

52 'Religious group' means a group of persons defined by reference to a religious belief or lack of religious belief.

(iii) The victim's membership (or presumed membership) of a sexual orientation group;[53]

(iv) A disability[54] or presumed disability of the victim.[55]

The legislation goes on to provide that if 'the offence is motivated (wholly or partly) by hostility' towards any of the above, the offence will also be one aggravated by hostility. Article 2(4) goes on to provide that it is immaterial whether the hostility is based to any extent on any other factor. The Order also makes provision for an increase in the maximum penalties available for certain offences, but these increases apply generally, and are not limited to cases which are aggravated by hostility, as was made clear in *R v Massey and Hawkins*.[56]

Jarman observes that police detection levels for hate crime in Northern Ireland have 'persistently' remained lower than for hate crime offences in other parts of the United Kingdom.[57] Further, Jarman notes that; when contrasted to comparable offences in Northern Ireland, detections have also remained persistently low.[58] When traced through the system, the attrition of the hate element of the offence is pronounced. As McVeigh articulates, things that were being labelled a 'hate crime' by the Police Service of Northern Ireland (PSNI) 'were not being processed as such by the criminal justice system'.[59] Jarman describes the attrition process:

> In the five years from 2007–2008 to 2011–2012 the PSNI recorded 13,655 hate incidents, including 9,376 hate crimes. These translated into 4.689 cases where the PSNI had gathered sufficient evidence to

53 'Sexual orientation group' means a group of persons defined by reference to sexual orientation. While transgender persons would not usually fall within this category, Criminal Justice Inspection Northern Ireland (CJINI) considers hate crimes against transgender persons to fall within the ambit of the legislation. See *Hate Crime: A Follow-up Inspection of Hate Crime by the Criminal Justice System in Northern Ireland* (CJINI 2010).

54 Disability for the purposes of the legislation means any physical or mental impairment, art 2(5).

55 Art 2, Criminal Justice (No 2) (Northern Ireland) Order 2004 (SI 2004/1991, NI 15)

56 [2008] NICC 2. These offences are, *inter alia*: malicious wounding or grievous bodily harm, increased from five years to seven years; assault occasioning actual bodily harm, increased from five years to seven years; common assault increased from 12 months and/or £5000 fine on summary conviction to maximum of two years and/or unlimited fine on conviction on indictment; criminal damage increased from 10 years maximum to 14 years maximum; putting in fear of violence increased from five years to seven years.

57 Neil Jarman, 'Acknowledgment, recognition and response: the criminal justice system and hate crime in Northern Ireland' in Haynes et al (n 7 above).

58 Ibid.

59 Robbie McVeigh, 'Hate and the state: Northern Ireland, sectarian violence and "perpetrator-less crime"' in Haynes et al (n 7 above) 406.

> pass a file to the PPS [Public Prosecution Service]. However, the PPS decided that in 2,743 of these cases, there was insufficient evidence for the crime to be considered 'aggravated by hostility'... This left 1,946 files that were potentially prosecutable as a 'hate crime', culminating in just 71 successful prosecutions ... Twelve cases were successfully prosecuted under [the 2004 Order] ... A hate crime recorded by the PSNI had less than a one per cent chance of resulting in a conviction aggravated by hostility.[60]

That said, the rates of recorded convictions have increased substantially since 2011–2012. For example, in its 2014–2015 Report the Public Prosecution Service shows that 53 defendants received an enhanced sentence under the 2004 Order,[61] and the 2015–2016 Report shows that, in 89 cases, an enhanced sentence was recorded where the aggravating element was proven.[62] Jarman speculates that this increase could be due to a number of factors: improvements in the quality of the evidence gathered; a greater awareness in the prosecution service and improvements in the preparation and presentation of cases; more effective 'joined up work' across the criminal process which allows cases to be tracked; and the increased attention given to the issue by the Department of Justice 'which holds the different criminal justice agencies to account'.[63] However, echoing Haynes and Schweppe's description of the 'disappearing' of hate crime in the current process in Ireland, McVeigh, argues that current legislation and policy are not effective across the criminal process:

> ... the legislation does not frame racist violence appropriately; the police do not police it appropriately; the PPS does not process it appropriately; the courts do not penalise it appropriately and the official statistics do not record it appropriately.[64]

The message that is thus being sent by the criminal justice agencies in Northern Ireland, McVeigh argues, is that while hate crime is a 'bad thing', it is not something which the criminal justice process is equipped to address. He concludes:

> Other criminal justice systems serious about addressing racist violence – including crucially the Republic of Ireland – should learn from the

60 Of the 71 prosecutions, the other 59 involved use of the Public Order (NI) Order 1987, the Protection from Harassment (NI) Order 1997 and the Criminal Attempts and Conspiracy (NI) Order 1983. Jarman (n 57 above) 61–62.

61 *Statistical Bulletin: Cases Involving Hate Crime 2014/2015* (Public Prosecution Service for Northern Ireland 2015).

62 *Statistical Bulletin: Cases Involving Hate Crime 2015/2016* (Public Prosecution Service for Northern Ireland 2016).

63 Jarman (n 57 above) 65.

64 McVeigh (n 59 above) 408.

> palpable failure of the Northern Ireland model. Northern Ireland provides a textbook example of how *not* to address hate crime.[65]

It is difficult to argue with McVeigh's argument that, to be effective, hate crime legislation must be both operational and effective: the enhanced sentencing model in Northern Ireland, he suggests, is incapable of being either.

The Independent Review of Hate Crime Legislation in Northern Ireland published its Consultation Paper, authored by Judge Desmond Marrinan, in January 2020.[66] The Review team found the statistics regarding the application of the 2004 order – and particularly that in 2018/2019, 'none of the 13 defendants received an increased sentence where the judge accepted that the aggravating feature ... had been proven beyond a reasonable doubt' – so troubling that they reviewed the transcripts of the 16 Crown Court cases referred to in statistics published by the Public Prosecution Service[67] in which the prosecutor considered the case to involve a hate crime aggravated by hostility. In fact, when the transcripts were analysed, the Review found that only four were prosecuted on the basis that the offence was aggravated by hostility: and in those four cases, the judges 'accepted the aggravating features' but either did not enhance the sentence, or, if they did, did not state that they were doing so.[68] This, the Review found, raised issues relating to the statistics published by the Public Prosecution Service, as well as how such cases are prosecuted and sentenced, though it was noted that these concerns were not new.[69]

Responses to the Consultation Process called for 'significant' changes in the law, and the introduction of aggravated offences, or for the introduction of the hybrid model, referred to by Marrinan as 'a statutory aggravation model'.[70] He recommends moving from an enhanced sentencing model to an aggravated offences one, going so far as to say 'that an aggravated offence model is the only means by which it can be consistently ensured that the hate element of a crime will be effectively addressed'.[71] This approach will, he suggests, 'have a much

65 Ibid 413 (original emphasis).

66 Independent Review of Hate Crime Legislation (n 3 above). For the purposes of transparency, it is to be noted that the author is a member of the Core Expert Group of the Review, though this group is advisory only.

67 *Statistical Bulletin: Cases Involving Hate Crime 2018/2019* (Public Prosecution Service for Northern Ireland Service 2019) 21.

68 Independent Review of Hate Crime Legislation (n 3 above) 59.

69 In particular, the review referred to *Racist Hate Crime: Human Rights and the Criminal Justice System in Northern Ireland* (Northern Ireland Human Rights Commission 2013).

70 Desmond Marrinan, *Hate Crime Legislation in Northern Ireland: Independent Review – Final Report Volume 1* (Department of Justice 2020) 113.

71 Ibid 125.

better chance of providing an effective approach' to dealing with hate crime, encouraging the police, as it will, to collect evidence at an early stage, and also ensure that the aggravation will be on the record of the defendant.

MODEL 3: SCOTLAND – THE 'HYBRID' MODEL

The Crime and Disorder Act 1998 also included legislation for addressing hate crime in Scotland, but utilised a very different model to that provided for in England and Wales. Section 96 of the 1998 Act is different to the aggravated offences created under sections 29–32 in that it can be applied to *any* offence. In this way, it is similar to an aggravated sentencing model. However, under the Scottish approach, the hate element is presented at the charge stage rather than at sentencing, which differentiates it from the aggravated sentencing model. Finally, and perhaps most importantly for the offender, like the aggravated offences model, where proven, the hate element is recorded on the criminal record of the offender. Table 2 illustrates the differences.

The legislation in place in Scotland applies where the section is specified in a complaint or labelled in an indictment, and where it is proved that the offence has been racially aggravated. In this model, the hate element of the offence is a sentencing matter, but is inserted into the case by way of a specific sentencing charge. Section 96(2) provides that an offence is racially aggravated for the purposes of the section if:

> (a) at the time of committing the offence, or immediately before or after doing so, the offender evinces towards the victim (if any) of the offence malice and ill-will based on the victim's membership (or presumed membership) of a racial group; or

	Aggravated offences	**Aggravated sentencing**	**Hybrid model**
Applies to any offence		**x**	**x**
Included at charge stage	**x**		**x**
Appears on criminal record of accused	**x**		**x**
Maximum penalty increased?	**x**		

Table 2: The Scottish 'hybrid' sentencing model

(b) the offence is motivated (wholly or partly) by malice and ill-will towards members of a racial group based on their membership of that group, and evidence from a single source shall be sufficient evidence to establish, for the purposes of this subsection, that an offence is racially aggravated.[72]

Where it is proved that the offence was racially aggravated, according to section 96(5), the court must:

(a) State on conviction that the offence was racially aggravated;

(b) Record the conviction in such a way that shows that the offence was racially aggravated;

(c) Take the aggravation into account when determining the appropriate sentence;

(d) And state what the sentence would have been if it was not so aggravated, and the extent or reasons for the difference, or the reasons for there being no such difference.

As is noted in the Bracadale Report, where an individual is convicted of an offence with a statutory aggravation, it will be recorded and taken into account at sentencing, will appear on the criminal record of the individual, and can be taken into account if the individual reoffends.[73] Bracadale was of the view that this statutory approach works well in practice: is extensively used; the approach has ensured that police and prosecutors are aware of the need to take a hate element into account; and it has facilitated the collation and publication of statistics.[74] That said, Chalmers and Leverick note that there is little reported case law on the provisions.[75] As well as the general provisions, Scottish law has created further aggravated offences: racially aggravated harassment under section 50A(1)(a) of the Criminal Law (Consolidation) (Scotland) Act 1995 and a further offence of racially aggravated behaviour created by section 50A(1)(b) of the Act.[76] It could be that the unusual approach taken to hate crime in Scotland is a product of the range of common law offences which still apply in that jurisdiction: it would be presumably

72 The provisions have now been extended to prejudice in relation to religion (s 74(2A) Criminal Justice (Scotland) Act 2003), disability (s 1(3) Offences (Aggravation by Prejudice) (Scotland) Act 2009), sexual orientation and transgender identity (s 2(3) Offences (Aggravation by Prejudice) (Scotland) Act 2009).

73 Alexander Campbell, Lord Bracadale, *Independent Review of Hate Crime Legislation in Scotland: Final Report* (Scottish Government 2018) (Bracadale Report).

74 Ibid.

75 James Chalmers and Fiona Leverick, *A Comparative Analysis of Hate Crime Legislation: A Report to the Hate Crime Legislation Review* (Scottish Government 2017).

76 See Bracadale Report (n 73 above) for a further analysis of these provisions.

impossible to create aggravated versions of these offences without equally making the base offence a statutory offence.[77]

From the review on statistics relating to hate crime conducted by the Crown Office and Procurator Fiscal Service, a number of observations can be made in relation to the operation of the Scottish legislation. This Report notes that, while racial hate crime is the most commonly reported hate crime, the number of charges reported is 33 per cent lower than the peak in 2011–2012.[78] Worryingly, 2018–2019 was in fact the lowest annual total since such figures were made available in 2003, and the first time that such figures went below 3000. The figures for 2019–2020 are only marginally higher. When we compare the numbers of charges of race crimes in relation to section 50A offences, and other charges made using the hybrid model, we see that, until 2014–2015, there were in fact *more* charges made under section 50A for aggravated harassment and behaviour than there were for all other offences prosecuted with a racial aggravation under the hybrid model. From 2014 to date, the trend is reversing, but there is still a large proportion of offences being prosecuted under the aggravated categories (see Table 3).

Though he does not refer to any statistics in relation to reported or recorded hate crime in Scotland, nor the number of prosecutions taken and the relative number of sentences imposed, Lord Bracadale states that he is 'satisfied that this approach has worked reasonably well', and he recommended that the approach be maintained.[79] However, in so recommending, the Bracadale report does not explore in any detail *how* the legislation is operating, or specify any indicators of effectiveness

Year	2010–11	2011–12	2012–13	2013–14	2014–15	2015–16	2016–17	2017–18	2018–19	2019–20
Total no of race crimes	**4178**	**4547**	**4034**	**4160**	**3820**	**3721**	**3367**	**3278**	**2921**	**3038**
Section 50A	2574	2792	2376	2300	1969	1757	1462	1370	1204	1208
Hybrid	1604	1755	1658	1860	1851	1964	1905	1908	1717	1830

Table 3: Crown Office and Procurator Fiscal Service hate crime statistics

77 In the Bracadale Report (ibid), it is noted that common law breach of the peace and common law threats are two of the most commonly charged offences in conjunction with statutory aggravations.

78 *Hate Crime in Scotland 2018–2019* (Crown Office and Procurator Fiscal Service 2019).

79 Bracadale Report (n 73 above) 14. Indeed, Lord Bracadale dedicates only three paragraphs to his analysis of the model for statutory aggravation in his 148-page report.

against which it might be considered. There is no analysis of the justice gap between the prevalence of hate crime, reports, prosecutions and convictions with respect to hate crime, nor any indications as to what those operating in the criminal justice system feel about the current regime.

Further, there is no analysis of the reasons for the relatively high rates of racially aggravated harassment and behaviour under section 50A(1) of the Act as compared to the figures relating to the hybrid model. This might simply be a product of the reported crimes, which would reflect police statistics for those offences. It might also be because those acting within the criminal justice process see the section 50A(1) offences as a more expedient, effective or pragmatic means of addressing hate crime when compared to the hybrid approach. This is particularly the case when we consider section 50A(1)(b), which has a similar non-aggravated offence in 'threatening and abusive behaviour'. Indeed, the fact that the usual rules requiring corroboration apply in relation to a prosecution under section 50A – that is, that more than one piece of evidence must be adduced to prove all parts of the offence – makes these figures even more difficult to understand: as the hybrid charge is a sentencing provision, only one piece of evidence is required under Scottish law, making it surely easier to prosecute. Indeed, Bracadale recommended the repeal of section 50A because existing legislation (and the utilisation of the hybrid charge) can fully address the mischiefs which section 50A seeks to address. Indeed, he refers to them in his Report as 'the two *alternative* routes'.[80] Further, in introducing the legislation, the then Lord Advocate noted that 'much of the behaviour which would be covered by the new standalone offence would also be covered by the crime of breach of the peace'.[81] While it may be the case that there are good reasons as to why prosecutors, in spite of the more onerous proof requirements, prosecute under section 50A, there are no reasons given, or any discussion had, as to the reasons for this.

PROPOSALS FOR LAW REFORM[82]

In considering models for law reform in England and Wales, whilst Walters et al note that the aggravated offences model was a useful one and, with some changes, could operate more effectively, they also note that creating a large number of aggravated offences risks 'bloating' the provision, would entail the creation of new statutory sentencing

80 Ibid 89 (emphasis added).

81 As cited in ibid 85.

82 At the time of writing, the Law Commission of England and Wales has yet to publish its consultation paper.

maxima and, ultimately, that some crimes would remain outside the aggravated offences model.[83] They recommend adopting the Scottish approach, which would aggravate any offence where there is sufficient evidence of hostility. This approach, they speculate, would address the concerns highlighted in their research, allowing juries and magistrates to determine whether an offence was in fact aggravated by hostility during the trial. They are clear, however, that, in order for this to operate as they suggest, it is essential that the aggravation 'make up part of the substantive offence which appears on the charge sheet'.[84] This, they state, is the only way to ensure that the hostility element forms part of the case and is addressed at trial as well as at sentencing.[85]

The Law Commission in its recent Consultation Paper considered this concern, as well as the issue that, by moving to a hybrid model, the increased maximum penalties would be removed. With respect to the first issue, the Commission observes that, while there are advantages to this approach, replacing an enhanced sentencing model with a hybrid model would mean that the hate element would have to be proven before a jury, resulting in aggravations being more difficult to secure and jury trials being longer. This issue, the Commission opined, was not present with respect to aggravated sentencing provisions. Removing the increased penalties, the Commission felt, would be undesirable, principally because of the negative message it would send. That said, this latter argument applies only where existing increased maxima exist.

In considering a model for Northern Ireland, Marrinan agrees that the Scottish approach would be a useful model to consider. His reasons for supporting this model are that: first, it applies to all protected characteristics; second, it can attach to any offence; and third, the aggravation is stated on conviction and accurately recorded. Agreeing with Walters et al,[86] he states that this approach would not require the extension of penalties across all offences, but rather that the legislation would provide that the sentence must be aggravated, and the extent of that uplift must be declared by the court.[87] Marrinan went on to propose that any new offence would include the following provisions where the offence was found to be aggravated by hostility:

83 Walters et al (n 39 above) 977.
84 Ibid. The authors do not provide any proposed wording for such a statutory provision.
85 Ibid.
86 Walters et al (n 32 above).
87 Independent Review of Hate Crime Legislation (n 3 above) 138.

- a requirement that the court 'shall' treat it as an aggravating factor;
- a requirement to state in open court that the offence was aggravated by hostility;
- a requirement that the conviction be recorded in such a way as to capture the aggravation;
- state the extent of and reasons for the difference in penalty which the court would have imposed if there were no such aggravation (and if there is no difference, the reason for there being no difference in cases of exceptional mitigation).[88]

Marrinan's recommendations thus are made on the basis of these criteria which he considers are what 'effective' hate crime legislation would look like. First, he clearly articulates the need for the equal treatment of protected characteristics, thus rejecting the approach in England and Wales and Scotland which creates a hierarchy of victims. Second, he suggests that the legislation should be applicable across a range of offences. Third, the legislation should facilitate and indeed require the 'message' of hate crime legislation to be delivered clearly – to the defendant in terms of the offence that is imposed and to society by stating that the offence was so aggravated in open court. Legislation should also facilitate the hate element of the crime being presented to the court, thus ensuring it is not 'disappeared' through the process; and it should ensure that the sentence is aggravated where a hate element is found to be present in the offence. Finally, the fact that an individual has been convicted of a hate crime should be recorded on the criminal record of the defendant to capture recidivistic behaviour.

In Ireland, the Scottish model was adapted as part of the (now lapsed) Private Members Criminal Justice (Aggravation by Prejudice) Bill 2016.[89] In responding to the provisions, with particular reference to the hybrid approach, the Minister for Justice had questions regarding how the proposal would operate in the context of a jury trial, and particularly observed that the Bill 'does not create an offence per se'.[90] This is, of course, true. The hybrid model, which facilitates the attachment of an aggravated sentencing provision onto any criminal offence by way of a specific charge, is not an offence which can be prosecuted independently, and does not have any specific penalty associated with it. Fundamentally, it is a sentencing provision. This point was also made by the South African Human Rights Commission in its consideration of the use of the model in the draft South African

88 Marrinan (n 70 above) 132.
89 Criminal Justice (Aggravation by Prejudice) Bill 2016.
90 Francis Fitzgerald, Dáil Debates Tuesday, 4 October 2016.

legislation.[91] Indeed Chalmers and Leverick note that the general rule in Scottish law is that an aggravating element – as opposed to an element of the offence itself – 'need not be proved by corroborated evidence',[92] and they frame the provisions as an example of a sentencing aggravation model.[93] In parliamentary debates on the Irish Bill, the Minister for Justice had a number of practical concerns regarding the proposal:

> When a trial involves a jury, it makes a determination on the offence with which the person is charged, such as assault. Under what circumstances would the motivation be determined? If it is not a matter for the jury, and I do not see how it can be, it seems difficult to envisage how this would operate in practice. It does not seem appropriate for a court following a verdict of guilty from a jury to state that the offence was aggravated in the manner set out in the Bill where this was not a matter determined by the jury.

Given the emphasis placed by Walters et al on the importance of the hostility element being dealt with at trial as well as sentencing in their proposed hybrid model,[94] these questions are particularly relevant and, as yet, unanswered. The hybrid model does not, in one fell swoop, create an aggravated version of all criminal offences, much as it might seek to act that way.

That said, if we accept Marrinan's criteria for what is expected of hate crime legislation, only the hybrid model is operational across all criteria. However, if we remove one – that is, the requirement that the legislation is operative across all offences – then the aggravated offences model is equally effective. Indeed, given the wealth of information we have in relation to the operation of that model in England and Wales as compared to the dearth of information we have in relation to the operation of the hybrid model, then it could easily be argued that the aggravated offences model is – on balance – more effective across all his criteria. Further, if we look to the operation of legislation in Scotland, we see that at least some of the racially aggravated offences under section 50A are prosecuted in preference to taking a prosecution for a simple offence using the hybrid aggravation. Despite Lord Bracadale's assertions, we cannot draw any conclusions as to the effectiveness of the operation of the model in Scotland in the absence of research. Indeed, the model seems unique at least across the Commonwealth,[95] and so there is no other means of considering its operation for the

91 SAHRC Submission to DOJCS regarding Hate Crimes and Hate Speech Bill (South African Human Rights Commission 2017).
92 Chalmers and Leverick (n 75 above) 9.
93 Ibid 44.
94 Walters et al (n 39 above) 977.
95 Goodall and Walters (n 3 above) 33.

purposes of law reform. Thus, we can only speculate that it might provide an appropriate and more effective response to hate crime than the aggravated offences model. Indeed, responses to the Bracadale consultation might suggest that aggravated offences (or what he calls 'standalone offences') are more appropriate than the hybrid model. The Coalition for Racial Equality and Rights, for example, emphasised the importance of the so-called standalone offence 'in conveying the serious nature and State condemnation of racial harassment'.[96]

Taking a cautious approach to law reform might lead us to a conclusion that the aggravated offences model is to be preferred. Indeed, Walters et al admit that the interests of defendants are best protected under the aggravated offences model[97] and note that interviewees in their study were of the view that it is more appropriate for defendants to be charged and prosecuted for an aggravated offence than have a judge determine the issue at sentencing, an issue which is particularly relevant in light of the Irish Minister for Justice's comments.[98] If the Scottish model is in fact simply an aggravated sentencing model which is different only because the hate element is in the charge, then one might well wonder if the issues highlighted by Walters et al[99] and Owusu-Bempah et al[100] which are associated with the sentencing model would apply in the context of the Scottish approach also. While recommending an aggravated offences model might be considered a conservative approach, I think that it is also a considered and evidence-based one. Thus, I favour a model of legislative reform which would create a range of aggravated offences, accompanied by a broader sentence enhancement model, which would operate in parallel to the aggravated offences model.

For these reasons, it is perhaps useful to consider the one outstanding criteria highlighted by Marrinan: that is, the applicability of the legislation across offences, which cannot be addressed using the aggravated offences model. If the aggravated offences model is preferred, the question as to which offences should be included in such legislation thus requires consideration. It can be argued that the list should include those types of offences most commonly perpetrated against the protected grounds identified in the legislation. We have seen in England and Wales, for example, that, while racist hate crime most commonly takes the form of offences against the person, criminal

96 Bracadale Report (n 73 above) 85.

97 Walters et al (n 39 above).

98 Abenaa Owusu-Bempah, Mark Walters, and Susann Wiedlitzka, 'Racially and religiously aggravated offences: 'God's gift to defence'?' (2019) 6 Criminal Law Review 463.

99 Walters et al (n 39 above).

100 Bempah et al (n 98 above).

damage and public order offences, in the context of disablist hate crime, fraud and forgery offences, robbery, burglary, and theft and handling offences are more common (see Table 4).[101]

If it is accepted that creating aggravated offences is the preferred approach, I argue that a range of aggravated offences should be introduced across all categories of protected grounds. The Law Commission sets out criteria for determining whether an aggravated version of an offence should be created, which are:

- the overall numbers and relative prevalence of hate crime offending as a proportion of an offence;
- the need to ensure consistency across the criminal law;
- the adequacy of the existing maximum penalty for the base offence; and
- whether the offence is of a type where the imposition of additional elements of the offence requiring proof before a jury may prove particularly burdensome.

While useful, these criteria cannot be considered exhaustive or determinative. If we assess the inclusion of aggravated forms of sexual offences, for example, as recommended by Walters et al, these might be considered particularly controversial, given the difficulties associated with the prosecution of such offences in the absence of an additional factor. Thus, according to the last criterion, these should be excluded. However, given that literature suggests that the crimes

Offence	**Racial and religious hate crime %**	**Homophobic /transphobic hate crime %**	**Disability hate crime %**
Homicide	0.06	0.00	0.55
Offences against person	76.45	59.23	48.01
Sexual offences	0.25	1.34	3.64
Burglary	0.34	0.28	8.72
Robbery	0.58	1.13	6.73
Theft and handling	1.69	1.69	12.14
Fraud and forgery	0.08	0.07	6.40
Criminal damage	3.27	3.94	2.87
Drugs offences	0.60	1.20	0.44
Public order offences	15.11	29.79	9.27

Table 4: Crown Prosecution Service data[102]

101 Fewer offences are listed as commonly attracting statutory aggravations in Scotland by Bracadale (n 73 above). The offences so listed are: common law breach of the peace; common law issuing threats; threatening or abusive behaviour; stalking; improper use of a public electronic communications network; and communicating indecently.

102 As set out in Walters et al (n 32 above) 57.

of sexual assault and corrective rape are particular manifestations of hate crime towards the LGBTQI community in particular, these acts should be capable of being recognised in any hate crime legislation.[103] Indeed, when we look across protected groups, theft and fraud offences are particularly prevalent with respect to disability hate crime, and, arguably, this would justify the creation of aggravated offences, though this might not be in compliance with the first criteria.

The approach that I advocate also allows the maximum sentence to be increased, to allow for a significantly higher sentence to be imposed, if that is what is desired on the part of lawmakers. Owusu-Bempah et al provide recommendations to enhance the operation of such offences, which could usefully be employed in both policing and training for criminal justice professionals. It is unquestionable that such legislation would be long and somewhat complex to draft and might be considered legislatively unwieldy, or 'bloated'. It is also unquestionable that the hybrid approach is more legislatively elegant and simple to construct. That said, simple statutory tools of construction might be used to alleviate some of this bloating, such as using a definitions section to define 'protected characteristic' to be used throughout, rather than create separate offences for each protected ground, and using a schedule to list the change in statutory maxima, if that is required. Ultimately, I would suggest that, while there is no evidence to suggest that the hybrid approach provides a statutory approach which is more effective than the aggravated offences model, there is ample evidence establishing that the aggravated offences model is effective when assessed against Marrinan's criteria.

CONCLUSION

It is generally accepted that, whilst more pragmatic, simple, and in keeping with the general operation of the criminal law, the enhanced sentencing approach to legislating against hate crime is not enormously effective. Further, as the Office for Democratic Institutions and Human Rights observes, 'a penalty enhancement, while easier to implement, may not fulfil the expressive function of recognizing and condemning a prohibited bias'.[104] From an operational perspective, we have seen in, for example, the Law Commission's recent Report in England and Wales, that the aggravated offences model produces a more effective

103 Laura C Hein and Kathleen M Scharer, 'Who cares if it is a hate crime? Lesbian, gay, bisexual, and transgender hate crimes – mental health implications and interventions' (2013) 49 Perspectives in Psychiatric Care 84.

104 *Hate Crime Laws: A Practical Guide* (Organisation for Security and Cooperation in Europe Office for Democratic Institutions and Human Rights 2009) 36.

response by the criminal justice process as compared to those offences in which the hate element is addressed only at sentencing.[105] Qualitative research conducted by Haynes et al in an Irish context indicates that the process most consistently recognises named offences: put simply, it was argued, addressing hate crime through sentencing provisions will not ensure that the hate element of a crime will be consistently addressed from the point of recording through to sentencing.[106] When compared to the enhanced sentencing model, the aggravated offences model has been shown to be more effective. Whilst theoretically, the 'hybrid model' has much to commend it, little analysis has been conducted on its operation and effectiveness, either in Scotland or elsewhere. While the question as to whether hate crime legislation should be introduced can be a politically sensitive one, the manner in which such legislation is framed is legally complex. A cautious and conservative approach to law reform is, I argue here, one which is most likely to be effective.

105 Law Commission (n 28 above).
106 Ibid.

Northern Ireland
Legal Quarterly
Autum Vol. 72 No. 3 (2021) 569–587
Commentaries and Notes DOI: 10.53386/nilq.v72i3.883

Motor insurers ignore the law, again

Dorothea Dowling*

Former Chair, Motor Insurance Advisory Board

Correspondence email: dowlingdorothea@gmail.com

ABSTRACT

At the 2020 AGM of the Chartered Insurance Institute it was conceded that COVID-19 had caused reputational damage to the sector because of its treatment of consumers during the current crisis. That, however, was at a time of public spotlight on only some of the underlying issues of corporate culture. An ongoing question that needs to be addressed is that of insurers' lack of fairness to claimants, particularly those injured by uninsured vehicles, and the failure or even refusal of governments to redress that imbalance. One aspect of that enquiry is addressed in this article.

Keywords: uninsured accidents; EU Directives; MIB; COVID-19; Brexit; contract; corrective justice.

INTRODUCTION

Trust in the insurance industry may be at an all-time low because of blanket refusals of claims by struggling businesses for financial losses following COVID-19, even where the contractual wording was crystal clear.[1] However, there are also other areas where the corporate culture of the insurance sector is in need of review, given its impact upon the corrective justice objective of tort.[2]

There are some motor accident victims in Northern Ireland who are being deprived of their compensation rights. If such declinatures were successfully challenged, those claimants could be entitled to

* Dr Dorothea Dowling is an independent consultant undertaking legal research on matters relating to insurance and is a recent PhD graduate of the University of Westminster, London.

1 In a test case by the Financial Conduct Authority, which was decided by the London Divisional Court in September 2020, most of the sample policy wordings were held to provide cover for business interruption losses following COVID-19 and the extent of coverage was actually extended by the UK Supreme Court in its decision of January 2021: *FCA v Arch Insurance (UK) Ltd & Ors* [2020] EWHC 2448 (Comm) and *FCA v Arch Insurance (UK) Ltd & Ors* [2021] UKSC 1.

2 S Hedley, 'The unacknowledged revolution in liability for negligence' in Sarah Worthington, Andrew Robertson and Graham Virgo (eds), *Revolution and Evolution in Private Law* (Hart 2018).

Francovich damages for the refusals by the Motor Insurers' Bureau (MIB) to comply with European Union (EU) law.[3]

Matters relating to uninsured vehicles were fundamentally changed by the 2017 decision of the Court of Justice of the EU (CJEU) that the MIB of Ireland (MIBI) is an emanation of the state. It was held that the Bureau has the responsibility for ensuring that EU Motor Insurance Directives are fully complied with to the benefit of claimants.[4]

A year later in England, the High Court held that the MIB in that jurisdiction was an emanation of the state.[5] That finding was upheld by the Court of Appeal in 2019.[6] On 13 February 2020 the UK Supreme Court refused an application by the MIB to appeal further.[7] This issue also has post-Brexit implications.[8]

The wording of many MIB agreements in Ireland and in the UK must now be considered defective to the point where they may be considered invalid in several contexts.

FIRST PRINCIPLES

The fundamental challenge that arises with MIB is its apparent perspective that it is merely 'a fund of last resort', resembling a benevolence system operated voluntarily by motor insurers.[9] While that might have been its role when the original 'gentleman's agreement' between insurers and the Government was signed in 1946, it is contestable whether that has been an accurate assertion since the First EU Motor Insurance Directive of 1972, and it is certainly not so since the Second Directive of 1984.

If the MIB were 'a fund of last resort' then claimants would be obliged to exhaust all other avenues of recovery first. In contrast, victims of defectively insured, uninsured, untraced, or unidentified vehicles can directly sue the MIB without first proving that the person responsible is unwilling or unable to pay. This is made clear in recitals to the Third Directive:

---

3 *Francovich v Italy* [1991] E R I-3061, [1995] ICR 722 and *Brasserie du Pecheur v Germany; R v SST ex p Factortame (No 4)* [1996] QB 405 ECJ.

4 The MIBI was first held in 1999 to be 'an emanation of the State' in the case of *Dublin Bus v MIBI* (29 October 1999) Dublin Circuit Court, Judge Bryan McMahon. This was subsequently confirmed by the 2017 decision of CJEU in *Farrell 2* – ECLI:EU:C:2017:745.

5 *Lewis v Tindale, MIB & Secretary of State for Transport* [2018] EWHC 2376.

6 *MIB v Lewis* [2019] EWCA Civ 909.

7 Lord Reed (President), Lady Arden and Lord Hamblen JJSC.

8 Under s 4 of the European Union (Withdrawal) Act 2018 a right under an EU Directive previously recognised in cases decided before exit day persists after Brexit.

9 Originally, the MIB granted *ex gratia* payments in 'hard cases' of severe injury but entirely at the unreviewable discretion of its board.

> Whereas, however, in the case of an accident caused by an uninsured vehicle, the victim is required in certain Member States to prove that the party liable is unable or refuses to pay compensation before he can claim on the body; whereas this body is better placed than the victim to bring an action against the party liable; whereas, therefore, this body should be *prevented from being able to require that the victim, if he is to be compensated, should establish that the party liable is unable or refuses to pay*.[10] (emphasis added)

In the absence of evidence of vehicle insurance, a claimant is exercising their rights under the EU Motor Insurance Directives and that trumps national law.

Such claims must be determined in a manner consistent with EU law as interpreted by the CJEU. The MIBs in Ireland and in the UK are obliged to handle those actions in accordance with established precedents on the assessment of damages in negligence at common law. There are very few exceptions permitted under the relevant EU Directives.

The Directives do not permit treatment by MIB that is less favourable to claimants than that which would apply to victims of other motor accidents where valid insurance is applicable to the vehicle. This is clear from the wording of the Sixth Directive of 2009 in the first paragraph of article 10 at chapter 4:

> Chapter 4 – Compensation for Damage Caused by an Unidentified Vehicle or a Vehicle for which the Insurance Obligation Provided for in Article 3 has not been Satisfied
>
> *Article 10*
>
> Body responsible for compensation
>
> 4 Each Member State shall apply its laws, regulations and administrative provisions to the payment of compensation by the body, without prejudice to any other practice which is *more favourable* to the victim. (emphasis added)

In this short commentary the focus is on just one category of claim where injured parties recover general damages for injuries but recovery for property damage losses such as vehicle repairs is refused because the culpable vehicle is deemed untraced. That approach by MIB is clearly not 'more favourable' than the law which applies to tort actions generally.

THE REAL ROLE OF THE MIB

The MIB is funded by law-abiding policyholders. It is not a voluntary levy on motorists, as it is hidden within the motor premium payable

10 The Third Directive became fully effective by 31 December 1995.

annually.[11] However, if they gave the matter any conscious thought, some consumers might consider it an additional cost worth paying lest they fall victim to an incident involving an uninsured driver.

The MIB is obliged to discharge a range of delegated responsibilities under EU Motor Insurance Directives. As summarised by the CJEU, these functions must be exercised in the public interest:

> Therefore, the task that a compensation body such as MIBI is required by a Member State to perform, a task that contributes to the general objective of victim protection pursued by the EU legislation relating to compulsory motor vehicle liability insurance, must be regarded as a task in the public interest that is inherent, in this case, in the obligation imposed on the Member States by Article 1(4) of the Second Directive.[12]

In contrast to the principles espoused in the passage above, where victim protection is the primary objective, my experience indicates that the priority of the MIB seems to be to curtail the exposure of insurers to the discharging of victims' legitimate rights under EU law.

There is an academic view that the courts may have been complicit in protecting the interests of insurers.[13] Alternatively, some judges have not been sufficiently alert to their own obligations to ensure that EU law rights are extended to claimants involved in motor vehicle accidents. Many of the existing precedents from the English and Irish courts must be treated with caution as being non-compliant with subsequent superior decisions of the CJEU.[14]

It is seriously arguable that the 2018 case of *RoadPeace* was wrongly decided to the extent that the starting point was the wording of the English Road Traffic Act 1988 rather than the EU Directives.[15] Additionally, the decision makes a number of references to the relevance of 'whether' the MIB is an emanation of the state when the reasoning of the CJEU in *Farrell 2* makes it clear that, if the MIBI in Ireland is an emanation of the state, the same must be so of the MIB in the UK.

Furthermore, the pre-condition to payment of property damage claims is currently the occurrence of a 'significant injury' which is defined as a minimum of four days' in-patient treatment, and this point was not determined in the *RoadPeace* case as is clear at paragraph 126

---

11 Data published in 2020 by the Central Bank of Ireland as insurance regulator indicates that approximately 3 per cent of motor insurers' income is paid to the MIBI.

12 Para 38 of *Farrell 2* (n 4 above).

13 Richard Lewis, 'Insurers' agreements not to enforce strict legal rights: bargaining with government and in the shadow of the law' (1985) 48 Modern Law Review 275.

14 Some of these now defunct precedents are examined in my article in the December 2016 edition of the *Gazette* of the Incorporated Law Society of Ireland, 'Defective motor insurance, EU law, and victims' rights'.

15 *R (RoadPeace Ltd) v Secretary of State for Transport* [2018] 1 WLR 1293.

of the decision. The reasoning offered for the amendment in England, Wales and Scotland of the 2011 MIB agreement to the lower threshold in the 2017 agreement, where the criteria included out-patient treatment, was to reflect 'changes in medical understanding, treatment and duration of hospital stay'. It would be difficult to argue that similar medical advances have not been made in Northern Ireland and, therefore, there should not be a divergence between jurisdictions on the trigger for payment in respect of property damage by using different definitions of 'significant injury'.

It is also important to highlight that at paragraph 133 of *RoadPeace* Ouseley J concluded:

> Certainly, whether the MIB is an emanation of the state may be a lively issue, but is *one to be pursued where an actual claim depends on it*. No point was taken in relation to the standing of RoadPeace to raise the issues which it has raised. But that does not mean that interesting issues, which probably have no practical application should be pursued by it, especially as *such issues can be pursued by affected litigants* when they do have practical application. (emphasis added)

If the *RoadPeace* case had acknowledged that the MIB was an emanation of the state, then the reasoning would have been different, and it might not have been left to future claimants to resolve these 'interesting issues' in further litigation.

When the MIB was found to be an emanation of the state in 2018 by the English High Court in *Lewis*,[16] it was held at paragraph 131 that 'the effect of European law was to treat the designated compensation body as if the obligation imposed on the State had been delegated to it in full'. That finding was not overturned on appeal, although Flaux LJ commented that matters could be resolved 'by amendment to the RTA and/or the MIB Articles of Association'. By the end of 2020, neither of those required steps had been taken. That indicates that the UK was knowingly in breach of the Directives while still a member of the EU.

THE IRRELEVANCE OF THE MIB 'GENTLEMAN'S AGREEMENTS'

The fundamental error into which the MIB appears to fall is reliance on wording of the Northern Ireland 2011 agreement while ignoring its obligations to ensure compliance with the EU Motor Insurance Directives as interpreted by the CJEU.

It is a matter for the discretion of the member states what national law will hold to be the extent of civil liability. In Ireland and the UK this rests on negligence principles at common law with some statutory

16 *Lewis v Tindale, MIB & Secretary of State for Transport* [2018] EWHC 2376.

interventions. However, it is the EU Directives which mandate the extent of compulsory insurance coverage necessary to deliver compensation to the full extent of that tort liability. This is made clear by article 3, as below:

> Each Member State shall, subject to Article 5, take all appropriate measures to ensure that civil liability in respect of the use of vehicles normally based in its territory is covered by insurance.
>
> The extent of the liability covered and the terms and conditions of the cover shall be determined *on the basis of the measures referred to in the first paragraph*. (emphasis added)

While domestic law on liability and domestic legislation on prosecutions for driving without insurance are within the competence of the member state, the extent of insurance cover is not.

In 2001 the MIB agreements were held not be provisions of national law according to the House of Lords.[17] That raises the question as to the standing of the contract between motor insurers and the Governments of Ireland and the UK as member states of the EU. In strict technical terms, injured parties have no privity to sue for damages on the basis of those agreements.[18] Those contracts were entered into for the benefit of claimants, but claimants would lack any effective power of enforcement.

Since these MIB agreements are determinable by notice from either of the two parties who are signatories, it may also be questionable whether a compensation body has been set up at all in compliance with article 10 which contemplates a permanent standing entity. That requirement seems not to be satisfied if the agreements are merely voluntary rather than being enforceable under the requirements of the Directives and relevant decisions of the CJEU.

The classification of these agreements as not being part of domestic law could indicate a failure by the member state to enact 'laws, regulations and administrative provisions' to ensure that compensation consistent with the Directives is provided to claimants. Clearly, the cleanest solution would have been primary legislation to provide clarity on the rights of accident victims.

The overarching objectives of these Directives are free movement and the harmonisation of compensation entitlements and claims procedures throughout the EU arising out of the use of motor vehicles, without limiting any national procedures which are more favourable to a claimant. This applies even where the motor insurance is defective and such claims must be handled by the compensation body (the MIB). This is reflected in article 10.4 of the Sixth (consolidating) Directive as below:

17 *White v White* [2001] UKHL 9, [2001] 1 WLR 481, [22].

18 *Bowes v MIBI* [2002] 2 IR79.

> Each Member State shall apply its laws, regulations and administrative provisions to the payment of compensation by the body, without prejudice to any other practice which is *more favourable* to the victim. (emphasis added)

There have been a number of enforcement actions by the European Commission against both Ireland and the UK where the transposition of EU law rights was held by the CJEU to be defective. For example, the failure by Ireland to correctly transpose the 1972 Motor Insurance Directive was found to be grounds for *Francovich* damages in *Farrell 2*.[19] It is relevant in the current context that liability for such damages rests with the insurance industry's MIB rather than being levied on the member state. It is not necessary to traverse the EU precedents here as the preciseness of the rights of claimants under the consolidating Sixth Directive means that these are currently directly effective rights.

The onus is upon national courts to give effect to the rights created by EU law, even where a variance exists as against national law and procedures. This has been reviewed in a number of CJEU precedents. The most frequently cited judgment is probably that of 13 November 1990 often referred to as *Marleasing*.[20] This case confirmed:

> that national courts must as far as possible interpret national law in the light of the wording and purpose of the Directive in order to achieve the result pursued by the Directive.

It was also held that national courts are required to:

> having regard to the usual methods of interpretation in its legal system, give precedence to the method which enables it to construe the national provision concerned in a manner consistent with the directive.

The onus rests upon the member state under article 5 of the Treaty of Rome to take all appropriate measures, whether general or particular, to ensure the fulfilment of that obligation. This binds all the authorities of member states including, for matters within their jurisdiction, the courts.

As held in *Farrell 2*, the CJEU highlighted that the EU principle of equivalence and effectiveness of remedies, even for the victims of uninsured or unidentified vehicles, requires that the latter must not face barriers to 'the general objective of victim protection ... in the public interest'. Therefore, it seems to follow that national legislation,

---

19 CJEU decision of October 2017 in *Farrell v Whitty & Ors* ECLI:EU:C:2017:745.

20 *Marleasing SA v La Comercial Internacional de Alimentacion SA*. Reference for a preliminary ruling: Juzgado de Primera Instancia e Instruccion no 1 de Oviedo – Spain. Directive 68/151/CEE – Article 11 – Consistent interpretation of national law. Case C-106/89. European Court Reports 1990 I-04135. ECLI identifier: ECLI:EU:C:1990:395.

such as the Road Traffic Acts and the litigation procedures, must be applied in a manner that is consistent with the Directives when it comes to victim compensation even if that is at variance with considerations under national law.

NON-COMPLIANCE BY THE UK AS A MEMBER STATE OF THE EU

Contrary to the objective of the EU Commission to harmonise claimant rights across the member states, the UK has introduced variances even within the geographical district of its own territory. There is a wide range of discrepancies, but only those relevant to the current focus are examined here.

The concept of 'significant injury' is not a term of art that acts as a threshold that must be overcome to commence a negligence action under English or Irish law. However, this barrier is defined in varying ways in the MIB agreements as summarised in Table 1.

The Isle of Man, which is not part of the UK nor of the EU, has an agreement with identical wording to that in Guernsey. Former overseas territories such as Hong Kong make no provision for victims

Table 1

Legal Jurisdiction	**Wording**
England, Scotland and Wales Dated 28 February 2017	Clause 7(2): The expression 'significant personal injury' in paragraph (1) means bodily injury resulting in— (a) death, or (b) 2 nights or more of hospital in-patient treatment, or (c) 3 sessions or more of hospital out-patient treatment.
Northern Ireland Dated 6 July 2011	Clause 3(a) amending 2004 Agreement: 'significant injury' means bodily injury resulting in death or for which 4 days or more of consecutive in-patient treatment was given in hospital, the treatment commencing within 30 days of the accident.
Gibraltar Dated 4 April 2003	Clause 5(1): This Agreement does not apply in the following cases—(a) where the applicant makes no claim for compensation in respect of death or bodily injury and the damage to property in respect of which compensation is claimed has been caused by, or has arisen out of, the use of an unidentified vehicle.
Guernsey Dated 26 September 2005	Clause 5(1): This Agreement does not apply where an application is made in any of the following circumstances … (a) where the applicant makes no claim for compensation in respect of death or bodily injury and the damage to property in respect of which compensation is claimed has been caused by, or has arisen out of, the use of an unidentified vehicle.
Jersey Dated 27 April 2005	Clause 5(1): This Agreement does not apply in the following cases—(a) where no death or bodily injury has been caused to any person and the damage to property in respect of which the application is made has been caused by, or has arisen out of, the use of an unidentified vehicle.

of unidentified vehicles, but they would not expect to be entitled to EU law rights.

Northern Ireland is not a separate member state of the EU and all citizens of the UK have identical EU law rights under the Motor Insurance Directives. There is no justification for the variances apparent in Table 1, nor is it permissible under EU law to accord more restricted rights to claimants in Northern Ireland. Claimants in Gibraltar, Guernsey and Jersey need only make a claim for personal injury to also recover for property damage caused by an unidentified vehicle. Indeed, in those jurisdictions both parts of such claims do not need to be made at the same time as two separate applications are permitted under those agreements.

In contrast, for claimants in Northern Ireland, the effectiveness of remedy is undermined for the victim of an unidentified vehicle by an unreasonable requirement of '4 days or more of consecutive in-patient treatment' because very few motor accidents result in such periods of hospitalisation. There is no provision under the Directives for more favourable treatment than in Northern Ireland to be accorded to citizens in England, Scotland and Wales where the threshold can be based on out-patient treatment as a trigger to also recover for property damage such as vehicle repairs.

Claimants in the Republic of Ireland are even more harshly treated in this context under the 2009 MIBI agreement as below:

> 7.1 The liability of MIBI for damage to property shall not extend to damage caused by an unidentified vehicle unless compensation for substantial personal injuries involving an inpatient hospital stay for five days or more has also been paid in respect of the event causing the damage subject to an excess of €500.

Research in the Republic of Ireland based on sample data indicates that for those who were hospitalised after road crashes 56.4 per cent had a length of stay of 1 to 2 days and an additional 15.5 per cent a stay of 3 to 4 days. Collectively, that represents 72 per cent of inpatients who would fall below the MIB hurdle before property damage could be recovered.[21]

As has been frequently stated in CJEU decisions, the aim of the Directives is:

> to guarantee that the victims of accidents caused by those vehicles receive comparable treatment *irrespective* of where in the European Union the accident occurred (see, inter alia, Case C-300/10 Marques Almeida [2012] ECR, paragraph 26 and the case-law cited).[22] (emphasis added)

21 'Admissions and costs to acute hospitals resulting from road traffic crashes' (*Irish Medical Journal,* 1 March 2012) .

22 This CJEU reference by Hungary related only to the First and Second Directives in the context of an insolvent mutual insurance company, but it explored the rationale of the Directives. Citation is from para 26 of *Csonka v Magyar* C-409-11 EU:C:2013:512 [2014] 1 CMLR14.

The wordings of the agreements in Table 1 are non-compliant with the Directives issued subsequent to 1984 in the context of unidentified vehicles. To quote from a CJEU decision in 2002, it is clear that such disparities should not exist:

> It is apparent from the third recital in the preamble thereto that the Second Directive was adopted in order to reduce the *major disparities* between the laws of the different Member States concerning the extent of the obligation of insurance cover. For that purpose, Article 1(1) and (2) of the Second Directive provides that the insurance referred to in Article 3(1) of the First Directive is to cover compulsorily both damage to property and personal injuries up to specific amounts.[23] (emphasis added)

These EU law rights of claimants have been progressively extended over time. The Second Council Directive of 30 December 1983 stated at article 1:

> 1 The insurance referred to in Article 3(1) of Directive 72/166/EEC shall cover compulsorily both damage to property and personal injuries.

However, at that time it then proceeded to provide an exception at paragraph 4 of article 1:

> 4 Member States may limit or exclude the payment of compensation by that body in the event of damage to property by an unidentified vehicle.

The consolidating Directive of 16 September 2009 highlights specific provision for damage caused by unidentified vehicles based on the rationale set out at preamble 14 as below:

> It is necessary to make provision for a body to guarantee that the victim will not remain without compensation where the *vehicle* which caused the accident is uninsured or *unidentified*. It is important to provide that the victim of such an accident should be able *to apply directly to that body as a first point of contact*. However, Member States should be given the possibility of applying *certain limited exclusions* as regards the payment of compensation by that body and of providing that compensation for damage to property caused by an unidentified vehicle may be limited or excluded in view of the danger of fraud. (emphasis added)

The 'danger of fraud' could arise where an owner causes damage to their own property and attempts to blame another vehicle when there was actually no such other party involved.

This points to a more fundamental challenge. The title of the MIB agreement refers to 'untraced drivers' whereas the Directives permit certain limited exclusions of property damage caused by 'unidentified vehicles'. This would be a classic 'hit and run' occurrence.

---

23 Para 4 of *Withers v MIBI* Case C-158/01.

PURPOSIVE APPROACH TO LIMITED EXCLUSIONS UNDER EU MOTOR INSURANCE DIRECTIVES

A further complication is introduced by the mischief of cloned vehicles widely covered in the media.[24] The sophistication of these schemes has also been revealed during criminal trials.[25] To quote from an Irish Court of Appeal decision on 2 March 2020, dismissing a challenge to sentencing:

> The charges to which the appellant pleaded guilty arise out of a professional and commercial transnational criminal conspiracy to clone motor vans stolen in the United Kingdom and sell them on in Ireland. The basic modus operandi involved motor vans being stolen and then being modified so as to change their chassis numbers and other identification numbers to match those of similar vehicles which had not been stolen, and copies of whose registration documents had been obtained by deception. This enabled the vehicles to be sold on with ostensibly valid documentation to unsuspecting purchasers in Ireland. The value of eight of the eleven motor vans in question ranged between Stg£9500 and Stg£18,864. There was no value available for three of the eleven motor vans.

These are well-organised scams. Obviously, right-hand drive vehicles most likely come from the geographical region of the British Isles. It is the responsibility of member states to ensure that all vehicles are insured and any failure in that obligation, challenging as it may be, should not prejudice claimants. To cite from *Csonka*:

> Article 3(1) of the First Directive – as has been pointed out in paragraph 28 above – requires each Member State, subject to the derogations allowed under Article 4 of that directive, to ensure that every owner or keeper of a vehicle normally based in its territory takes out a policy with an insurance company for the purpose of covering, up to the limits established by European Union law, his civil liability arising as a result of that vehicle. Viewed in that light, the very fact that damage has been caused by an uninsured vehicle attests to a breakdown in the system which the Member State was required to establish and justifies the payment of compensation by a national body providing compensation.[26]

The restrictions noted above on property damage claims by victims of uninsured drivers are difficult to reconcile with this opinion.

The extent of uninsured driving in the EU is well known and in some jurisdictions is growing.[27] As the CJEU recorded at paragraph 39 of its

---

24 As just one example: Rob Hull, 'Capital clones: how London is enduring a crime wave of cars with cloned plates creating havoc for owners and councils' (*This is Money*, 21 August 2019).

25 *Director of Public Prosecutions v Reilly* [2020] IECA 47.

26 *Csonka* (n 22 above) para 31 extract.

27 European Motor Insurance Markets Report, February 2019.

judgment in *Farrell* 'the intervention of (the Art 10 body) is designed to remedy the failure of a Member State to fulfil its obligation to ensure that civil liability in respect to the use of motor vehicles normally based in its territory is covered by insurance'.

It seems that, in the context of claimants in Northern Ireland, the UK is over-exploiting the limited restrictions on compensation provided for in the Third Motor Insurance Directive at preamble number 17, as below:

> The option of limiting or excluding legitimate compensation for victims on the basis that the vehicle is unidentified should not apply where the body has paid compensation for significant personal injuries to any victim of the accident in which damage to property was caused. Member States may provide for an excess, up to the limit prescribed in this Directive, to be borne by the victim of the damage to property. The conditions in which personal injuries are to be considered significant should be determined by the national legislation or administrative provisions of the Member State where the accident takes place. In establishing those conditions, the Member State may take into account, inter alia, whether the injury has required hospital care.

The term '*inter alia*' in the last sentence of the above passage is also important because it indicates a discretion and implies that other factors may also be taken into account.[28] The danger of fraud should not be assumed in every individual case.[29]

In the context of cloned vehicles, disputes about responsibility for lack of insurance can be informed by the Fourth Directive, which provides some guidance as to the designated jurisdiction. This 2000 Directive introduced the direct right of action in cross-border accidents provided in article 7 as below:

> If it is impossible to identify the vehicle or if, within two months following the accident, it is impossible to identify the insurance undertaking, the injured party may apply for compensation from the compensation body in the Member State where he resides. The compensation shall be provided in accordance with the provisions of Article 1 of Directive 84/5/EEC. The compensation body shall then have a claim, on the conditions laid down in Article 6(2) of this Directive:
>
> (a) where the insurance undertaking cannot be identified: against the guarantee fund provided for in Article 1(4) of

28 The French language version of the Directive expresses the last two words in the passage above as *soins hospitaliers* which also translates to hospital care but in neither language does this equate to the higher barrier of in-patient treatment of four or more days.

29 Despite the relatively low number of prosecutions pursued, the federations of insurers across Europe collectively adopt the position that 10 per cent of all claim payments involve fraud. See 'Insurance fraud: not a victimless crime', November 2019.

> Directive 84/5/EEC in the Member State where the vehicle is normally based;
>
> (b) in the case of an unidentified vehicle: against the guarantee fund in the Member State in which the accident took place;
>
> (c) in the case of third-country vehicles: against the guarantee fund of the Member State in which the accident took place.

In cross-border accidents where there is doubt, then the compensation fund in the jurisdiction where the vehicle is normally based is mandated to deal with the claim. The Fifth Directive of 2005 extended the provisions of the Fourth Directive to all accidents in the EU.

If a cloned vehicle was displaying Northern Ireland registration plates it seems likely, on the balance of probabilities, that it was based in Northern Ireland.

Cloned vehicles must be distinguished from those involved in a 'hit and run' accident because they are not unidentified but rather are uninsured and resultant claims should be handled on those terms, which provide compensation for both injury and property damage. Even the guidelines to the Northern Ireland MIB agreement state at paragraph 11.2 that it can be unclear in some instances whether the Untraced Drivers Agreement or the Uninsured Drivers Agreement applies, and it advises claimants to make applications under both agreements. In any such cases of doubt the provisions which are more favourable to the claimant should apply in accordance with EU law. I have encountered untenable arguments by the MIB that a car bearing false registration plates was untraced and outside the terms of the agreement when that vehicle was still stuck into the side of a bus as recorded in photographic evidence.

Two further points can be highlighted at this stage.

It will be noted that the headings employed by MIB refer to 'Drivers' Agreement whereas the EU Directives focus on the vehicle and are silent on the traceability of the driver. This is an irreconcilable variance from the approach in mainland Europe because Ireland and the UK focus on driver use in legislation on compulsory motor insurance.

Secondly, in the wording of the Third Directive at Preamble number 17 cited above, it may be noted that the provisions are to be introduced in 'national legislation or administrative provisions of the Member State'. Again it is questionable whether the 'gentleman's agreement' with the MIB can be assigned such status within the canons of the separation of powers, and concerns about excessive delegation of legislative functions may arise.[30]

---

30 Bruce Carolan, 'Separation of powers and administrative government' in E Carolan and O'Doyle (ed), *The Irish Constitution: Governance and Values* (Thomson Round Hall 2008) 225.

UNWARRANTED RESTRICTIONS ON EFFECTIVE REDRESS

Claims against the MIB are standard civil actions for damages. It was necessary for that point to be clarified by the Irish High Court in the context of authorisation by the Personal Injuries Assessment Board before litigation could be commenced.[31] Relevant extracts from that reasoned decision are cited below:

> It is a canon of construction that words are primarily to be construed in their ordinary meaning or common or popular sense and as generally understood unless the context requires some special or particular meaning to be given: *Stephens v Cuckfield* R.D.C. 1962 All ER 716 at 719. Halsburys Laws of England Third Edition Volume 1 at paragraph 9 has this to say –
>
> 'The popular meaning of the expression "cause of action" is that particular act on the part of the Defendant which gives the Plaintiff his cause of action.'
>
> In this sense the cause of action against the uninsured Defendants is negligence. As against the Bureau there is strictly speaking no cause of action as the law does not confer upon a non-party a right to sue upon a contract: Bowes v Motor Insurers Bureau of Ireland supra. However it is the negligence of the uninsured Defendants that triggers the proceedings and without which act the proceedings could not be maintained even with the concession invariably made by the Bureau that an action is maintainable against it. For this reason I take the view that the cause of action against the Bureau is the same as that against the uninsured Defendants. While the relief claimed may be a declaration or specific performance in terms of pleading, the intention of joining the Bureau in an action is to recover damages for negligence awarded against uninsured Defendants.[32]

The reason that it is important to acknowledge that these are negligence actions for damages against the MIB is because MIB claimants are dealt with less favourably than other such tort plaintiffs. This is questionable relative to the equivalence rights of claimants under EU law, without prejudice to more favourable treatment, and may also offend the principles of natural justice and fair procedures.

As reflected in the wordings of the agreements, the MIB is operating in a quasi-judicial manner without any legislative base to ground its procedures.[33] The MIB purports to make binding findings of fact which

---

31 *Campbell v O'Donnell & Ors* [2005] IEHC 266.

32 Paragraphs are not numbered but this appears at page 12 of the 14-page decision.

33 Cl 11(3).

impact on final determination of rights under domestic law quite apart from any EU dimension.[34]

The only redress available to dissatisfied claimants under the terms of the UK agreements is imposed arbitration.[35] That creates a number of difficulties.

Under the terms of the Consumer Rights Act 2015, an arbitration clause could be deemed unfair.[36] That Act at schedule 2 provides a list of terms that may be regarded as unfair under the categories below:

> 20 A term which has the object or effect of excluding or hindering the consumer's right to take legal action or exercise any other legal remedy, in particular by—
>
> (a) requiring the consumer to take disputes exclusively to arbitration not covered by legal provisions,
>
> (b) unduly restricting the evidence available to the consumer, or
>
> (c) imposing on the consumer a burden of proof which, according to the applicable law, should lie with another party to the contract.

The MIB also restricts the evidence available to consumers in Northern Ireland because it makes only a fixed contribution to legal costs rather than indemnifying at the rate which would apply to negligence actions.[37] The imbalance of power between the parties is then further stacked against the claimant who is hampered in seeking an expert opinion because any disbursements must be authorised in advance by MIB.[38]

Aside from these real financial and procedural burdens placed on MIB claimants, the difficulty with being deprived of a court adjudication of rights is that arbitration outcomes are confidential to the parties, so there is no public access to the reasoning.[39] This adjudication process prevents a publicly available jurisprudence emerging from which parties might better understand their rights. It also conceals from regulators the extent to which injured parties are being denied their rights when relying on financial services for protection.

---

34 Cl 12(2).

35 Cl 15.

36 EU (then EEC) Unfair Consumer Contract Terms Directive 93/13/EEC, implemented in domestic law by Unfair Terms in Consumer Contracts Regulations 1999 (SI 1999/2083) and replaced by the Consumer Rights Act 2015.

37 Cl 21 (contribution towards legal costs).

38 Cl 21(11)(a).

39 In the Irish MIB Agreement arbitration is only imposed (while still objectionable) to disputes about compliance with pre-conditions at cl 3 on notification procedures which are purportedly conditions precedent to liability.

The argument that MIB is providing a financial service finds some support from a number of sources. Of the four issues to be tackled in the next Motor Insurance Directive, two are specifically assigned to the Consumer Financial Services Plan.[40] Those two include the 'role and functioning of motor guarantee funds'. The Sixth Directive also places emphasis on enhancing the single market for motor insurance and has as an objective the removal of discrepancies which could present a barrier to cross-border services. A corollary of that assertion is that the MIB should be bound by Codes of Conduct issued by the Financial Conduct Authority. An additional benefit of that approach is that disputes in MIB claims would be referable by consumers to the Financial Services Ombudsman process which does not levy any fees for its investigation and adjudication services.

This is not just a theoretical matter of interest to the academic. Even post-Brexit it will be necessary to replace previous EU protection of claimants with mechanisms in domestic law that respect natural justice and fair procedures.

CONCLUSION

In the MIB Untraced Drivers Agreement, there is only one reference to EU Directives and that is at the penultimate page of a 42-page document. Compliance with the Directive is claimed as follows:

> This Agreement, being made for the purposes of Article 10 of the Consolidated Motor Insurance Directive 2009/103/EC of 16 September 2009—

As detailed in this article, that agreement is clearly non-compliant and knowingly so given the decisions of the CJEU.

In the 13-page MIB Uninsured Drivers Agreement, there is not a single reference to EU Motor Insurance Directives. Indeed, many clauses are clearly non-compliant and, again, knowingly so given the decisions of the CJEU.

As a further example of non-compliance, the wording of the agreement purports to limit the MIB liability to unsatisfied judgments as below:

> **MIB's obligation to satisfy claims**
>
> 3(1) Subject to the exceptions, limitation and preconditions set out in this Agreement, if a claimant has obtained an unsatisfied judgment against any person in a Court in Great Britain then MIB will pay the relevant sum to the claimant or will cause the same to be so paid.

40 EU Consumer Financial Services Plan (COM (2017)) 139.

That divergence with EU law has wide-ranging implications that extend beyond the scope of the succinct issue of recovery for property damage caused by unidentified vehicles and will be the subject of a subsequent article.

Placing the onus on claimants to secure judgments against culpable parties is directly contrary to the Third Directive. Placing 'exceptions, limitation and preconditions' on claimants is an unjustified attempt by a mere gentleman's agreement to subordinate the EU law reflected in CJEU decisions on the interpretations of the Motor Insurance Directives.

Citizens of the EU are entitled to harmonised rights. There is certainly no justification for the variances identified in this article between the rights of Northern Ireland claimants as compared with those in England, Scotland and Wales, nor for the gap between the EU provisions and domestic law and practice.

Of course, now that the UK has ceased its membership of the EU it can remove rights that citizens have under the Directives and CJEU decisions, some of which have been reviewed in this article. However, such reduced entitlements will only apply to motor accidents from some future date upon which such national legislation is commenced.

One step in that direction has been taken by the publication on 21 June 2021 of a Private Members' Bill to amend retained EU law.[41] The objective is to roll back on the 2014 CJEU decision in *Vnuk* where the accident involved a tractor on a farm and that interpretation of article 3(1) of the First Directive (71/166/EEC) extended compensation rights to victims of occurrences on private lands to which the public had access.[42] The amending Bill reached second reading in the House of Commons on 29 October 2021 and has now been sent to a Public Bill Committee. The proposed mechanism is to insert a new section 156A after section 156 of the Road Traffic Act 1988 to confine compulsory insurance requirements to public roads for what might be regarded as the traditional definition of motor vehicles. The current draft wording is as below:

156A Retained EU law relating to compulsory insurance

(1) To the extent that Article 3 of the 2009 Motor Insurance Directive (as it had effect at any time) is relevant to any question as to the interpretation or effect of any provision of this Part, references in that Article to liability in respect of the use of vehicles are to be read as not including liability in respect of the use in Great Britain of vehicles—

41 Motor Vehicles (Compulsory Insurance) Bill 2021.

42 *Damijan Vnuk v Zavarovalnica Trigalev* (C-162/13).

(a) other than motor vehicles, or

(b) otherwise than on a road or other public place.

(2) Subsection (1) does not apply in relation to any question for the purposes of section 145(3)(aa) or (b) as to the interpretation or effect of the law on compulsory insurance of, or applicable in, a Member State or Northern Ireland.

(3) Relevant section 4 rights cease to be recognised and available so far as they relate to compensation in connection with the use in Great Britain of vehicles—

(a) other than motor vehicles, or

(b) otherwise than on a road or other public place.

(4) Accordingly, to the extent that it is inconsistent with subsection (1) or (3), retained case law ceases to have effect.

(5) In this section—

"the 2009 Motor Insurance Directive" means Directive 2009/103/EC of the European Parliament and of the Council of 16 September 2009 relating to insurance against civil liability in respect of the use of motor vehicles, and the enforcement of the obligation to insure against such liability;

"relevant section 4 rights" means section 4 rights which—

(a) are recognised and available in the law of England and Wales or the law of Scotland, and

(b) derive from the obligation imposed on the United Kingdom by Article 10 of the 2009 Motor Insurance Directive as it had effect immediately before IP completion day (which relates to compensation in connection with the use of vehicles in cases where drivers are uninsured or untraced);

"retained case law" has the same meaning as in the European Union (Withdrawal) Act 2018 (see section 6(7) of that Act);

"section 4 rights" means rights, powers, liabilities, obligations, restrictions, remedies and procedures which continue to be recognised and available in domestic law by virtue of section 4 of the European Union (Withdrawal) Act 2018 (saving for rights etc under section 2(1) of the ECA), including those rights, powers, liabilities, obligations, restrictions, remedies and procedures—

(a) as modified by domestic law from time to time, and

(b) as they apply to the Crown.

> (6) Nothing in this section applies in relation to the use of a vehicle before the day on which section 1 of the Motor Vehicles (Compulsory Insurance) Act 2021 comes into force.[43]

Two points warrant emphasis. First, the very existence of this Bill puts beyond doubt that rights secured by claimants under EU law still apply to motor accidents in the post-Brexit era until such time as national legislation alters that position. Secondly, it will be noticed that this Bill does not apply to Northern Ireland and such 'jurisdictional' variances are only possible after the end of the transition period because the UK was a single Member State bound throughout by the provisions of the 2009 and previous Directives as interpreted by the CJEU. Legal practitioners will need to be mindful of the litigation minefield that is likely to be created by the parallel frameworks as between national and EU law and as between Northern Ireland compared to England, Scotland and Wales.

43 Motor Vehicles Bill (n 41 above).

Northern Ireland
Legal Quarterly
Autumn Vol. 72 No.3 (2021): 588–595
Commentaries and Notes DOI: 10.53386/nilq.v72i3.981

Surrogacy and public policy*

Elaine O'Callaghan†

University College Cork

Correspondence email: elaine.ocallaghan@ucc.ie

ABSTRACT

The Supreme Court in the United Kingdom has held that it is not contrary to public policy to award damages in tort to fund a commercial surrogacy in another jurisdiction where this is lawful. This significant decision, in the case of *Whittington Hospital NHS Trust v XX* [2020] UKSC 14, will potentially have an impact on the regulation and reform of surrogacy law in the United Kingdom, Ireland and internationally. The judgment delivered by Lady Hale draws attention to multiple inconsistencies in the law, and it highlights, in particular, the need for effective regulation of domestic surrogacy. Legislators face an important and imminent challenge to reconcile the reality of commercial surrogacy with a deficient legal framework. This article seeks to highlight some of the important issues which this case has raised when considering regulation and reform of surrogacy law.

Keywords: commercial surrogacy; public policy; law reform; domestic surrogacy; international surrogacy.

INTRODUCTION

Proposals for the regulation and reform of surrogacy law have been published in recent years in Ireland and the United Kingdom (UK).[1] In Ireland, where surrogacy is currently unregulated, the General Scheme of the Assisted Human Reproduction Bill 2017 (AHR Bill 2017) provides for domestic, gestational and non-commercial, or altruistic, surrogacy.[2] No provision is made for the regulation of international surrogacy arrangements, while commercial surrogacy is expressly prohibited under Head 40 of the AHR Bill 2017.[3] A surrogate can, however, claim 'reasonable expenses'.[4] This proposed approach

* First published in NILQ 72.AD2 (2021) ADVANCE 2 August 29–35.

† I would like to thank the anonymous reviewers for their valuable comments on an earlier draft of this piece. Any errors and omissions remain my own.

1 In Ireland, see the General Scheme of the Assisted Human Reproduction Bill 2017 (Department of Health 2017). In the UK, see the Law Commission of England and Wales and Scottish Law Commission, *Building Families through Surrogacy: A New Law. A Joint Consultation Paper* (Law Com No 244, 2019).

2 Head 36 of the AHR Bill 2017.

3 Head 40(2) outlines that commercial surrogacy involves payment and/or reward in respect of a surrogacy agreement.

4 Head 41 of the AHR Bill 2017.

prohibiting commercial surrogacy arrangements is in line with how many jurisdictions have chosen to regulate surrogacy.[5] According to the UN Special Rapporteur on the sale and sexual exploitation of children, including child prostitution, child pornography and any other child sexual abuse material (hereinafter UN Special Rapporteur), this is 'based on the viewpoint that commercial surrogacy commonly commodifies children and exploits surrogate mothers'.[6]

Commercial surrogacy is also prohibited in the UK, although this position has been queried by many, in part because of difficulties expressed by judges in case law such as *Re X and Y (Foreign Surrogacy)*.[7] As the Law Commission of England and Wales and the Scottish Law Commission (the Law Commissions) have stated, 'the current law, which enables surrogates to be paid "expenses reasonably incurred" has been interpreted widely'.[8] Jackson commented that 'in practice, UK citizens can engage in commercial surrogacy abroad or at home without facing any sanction at all. The prohibition on commercial surrogacy is therefore almost completely ineffective.'[9] Fenton-Glynn and Scherpe also emphasise this point, stating that, '[t]he law is not being enforced, and commercial agreements are being permitted through the back door. This undermines the rule of law by allowing a practice that the legislature has expressly disallowed'.[10] The Law Commissions suggested that using the terms 'altruistic' and 'commercial' in the reform of surrogacy law can be unhelpful, thereby also recognising the many discrepancies that permeate law and practice.[11]

---

5 Hague Conference on Private International Law, *A Preliminary Report on the Issues arising from International Surrogacy Arrangements* (March 2012) 13, para 18.

6 Report of the UN Special Rapporteur on the sale and sexual exploitation of children, Thematic Report on Surrogacy, A/HRC/37/60 (15 January 2018) 5, para 15 and at 7, para 20.

7 Section 54(8) Human Fertilisation and Embryology Act 2008. *Re X and Y (Foreign Surrogacy)* [2008] EWHC 3030 (Fam), [2009] 1 FLR 733. For a discussion of this case and wider reform, see C Fenton-Glynn, 'Outsourcing ethical dilemmas: regulating international surrogacy arrangements' (2016) 24(1) Medical Law Review 59.

8 Law Commission of England and Wales and Scottish Law Commission (n 1 above) 26, [2.15].

9 E Jackson, 'UK Law and International Commercial Surrogacy: "the very antithesis of sensible"' (2016) 4(3) Journal of Medical Law and Ethics 197, 203.

10 C Fenton-Glynn and J Scherpe, 'Surrogacy: is the law governing surrogacy keeping pace with social change?' (Cambridge Family Law 2017) 4.

11 The Law Commission defined commercial surrogacy as: '[a] surrogacy arrangement in which the woman who becomes the surrogate and any agency involved charge the intended parents a fee which includes an element of profit. A commercial surrogacy arrangement may also be characterised by the existence of an enforceable surrogacy contract between the intended parents and the surrogate'. Law Commission of England and Wales and Scottish Law Commission (n 1 above) xiv.

In this context, the judgment of the Supreme Court in the UK in the case of *Whittington Hospital NHS Trust v XX* presents an important, and likely an influential, view of the stance of international commercial surrogacy arrangements.[12] Lady Hale held that it is possible to get damages in tort for a commercial surrogacy arrangement carried out abroad. The next section of this article presents an outline of this case, detailing the issues which arose before the court and how Lady Hale navigated them, especially in overturning the decision of the Court of Appeal in *Briody v St Helen's and Knowsley Area Health Authority* (*Briody*).[13] The dissenting views of Lord Carnwath and Lord Reed are also presented. The third section presents a discussion of several important issues which arose in the judgments. It is argued here that these issues must be considered by legislators for the regulation and reform of surrogacy law. The final section sets out the conclusions.

WHITTINGTON HOSPITAL NHS TRUST V XX

In this case, Lady Hale delivered the judgment of the Supreme Court, with Lord Kerr and Lord Wilson concurring. As to the facts, the claimant's cervical smear tests and biopsies were wrongly reported on multiple occasions. The errors came to light in 2013, some five years after the claimant's initial wrongly reported smear test and, at that stage, her condition meant that she could no longer bear a child. In advance of the claimant's surgery and chemo-radiotherapy, she underwent treatment to collect and freeze eight eggs. The case before the Supreme Court considered what damages can be recovered for losing the ability to bear a child and focused on three issues:

(1) Are damages to fund surrogacy arrangements using the claimant's own eggs recoverable?

(2) If so, are damages to fund surrogacy arrangements using donor eggs recoverable?

12 *Whittington Hospital NHS Trust v XX* [2020] UKSC 14.

13 [2001] EWCA Civ 1010, [2002] 2 WLR 394. A number of case comments have already been published regarding this judgment. See, for example, K Horsey and A Powell, 'A step too far? *Whittington Hospital NHS Trust v XX* [2020] UKSC 14' (2021) 29(1) Medical Law Review 172; N Bhatia, 'Whittington Hospital NHS Trust v XX [2020] UKSC 14' (2020) 17 Bioethical Inquiry 455–460. For analysis of this case at the Court of Appeal, see: J L M Taylor, 'International commercial surrogacy as a new head of tortious damage: XX v Whittington Hospital NHS Trust [2018] EWCA Civ 2832' (2020) 28(1) Medical Law Review 197. See also, B M Dickens, 'Paid surrogacy abroad does not violate public policy: UK Supreme Court' (2020) 150(1) International Journal of Gynaecology and Obstetrics 129.

> (3) In either event, are damages to fund the cost of commercial surrogacy arrangements in a country where this is not unlawful recoverable?[14]

The claimant stated that she wanted to have four children. While she had frozen eight eggs, evidence before the court suggested that it was likely that she could have two children using her eggs and donor eggs would be required to have more children. The claimant expressed preference to use commercial surrogacy in California. The High Court, in following the decision of the Court of Appeal in *Briody*,[15] held that commercial surrogacy in California was 'contrary to public policy' and that 'surrogacy using donor eggs was not restorative of the claimant's fertility'.[16] Non-commercial surrogacy using the claimant's own eggs was restorative of her fertility.[17] The claimant appealed the decision and the Court of Appeal held that '(p)ublic policy was not fixed in time ... Attitudes to commercial surrogacy had changed since Briody; perceptions of the family had also changed and using donor eggs could now be regarded as restorative.'[18] The hospital appealed this decision, bringing it before the Supreme Court.

Lady Hale held that the Supreme Court was not bound by the *ratio* of *Briody*. She held that 'developments in law and social attitudes' as well as 'useful information' documented in the Law Commissions' Consultation Paper demonstrate that a different conclusion must be reached in this case.[19] Lady Hale described the changes that have occurred since *Briody*, including that there are now a number of organisations in the UK providing not-for-profit surrogacy arrangements, a wider recognition of 'the family', government recognition that surrogacy is a pathway to parenthood and medical developments. Lady Hale also cited the Law Commissions' summary of social attitudes in this area, namely that:

> ... the research that exists suggests that public attitudes to surrogacy also now stand in stark contrast to the prevailing hostile attitudes at the time of the [Surrogacy Arrangements Act] 1985. The available research reflects the fact that the legislation is now out of step with attitudes towards surrogacy.[20]

In making a decision, Lady Hale stated that:

---

14 *Whittington Hospital* (n 12 above) [8].
15 [2001] EWCA Civ 1010, [2002] 2 WLR 394.
16 *Whittington Hospital* (n 12 above) [6].
17 Ibid [6].
18 Ibid [7].
19 Ibid [28].
20 Ibid [37].

> ... (n)othing which the claimant proposes to do involves a criminal offence either here or abroad. Her preferred solution is a Californian surrogacy which is lawful there and UK law does not prohibit her from arranging or taking part in it.[21]

On the first issue before the Supreme Court, Lady Hale held that it is possible to claim damages for the cost of surrogacy using the claimant's own eggs. On this matter, she cited the 'acceptance and widespread use of assisted reproduction techniques, for which damages are payable' and reiterated Sir Nelson's decision in the High Court wherein he cited the *Briody* case: given the right evidence of the reasonableness of the procedure and the prospects of success, such a case should be capable of attracting an award.[22]

The second issue to be determined was whether damages are recoverable to fund surrogacy using donor eggs. Counsel for the claimant stated 'that this is no different from other artificial means of replacing what has been lost, for example, by having an artificial limb fitted to replace the one which has been amputated'.[23] This argument, coupled with the changing definition of the family in the years since *Briody*, was persuasive for Lady Hale and she held that 'subject to reasonable prospects of success, damages can be claimed for the reasonable costs of UK surrogacy using donor eggs'.[24]

The final issue to be addressed by the court was whether damages are recoverable for a commercial surrogacy in a country where it is lawful. While the judge acknowledged that surrogacy contracts are not enforceable in the UK, she explained that many of the expenses involved in a commercial surrogacy in California would also be payable in the UK. Further, she stated that the deterrent in UK legislation that a court may refuse to retrospectively authorise payments does not appear to have been done in practice, given the court's consistent focus on the child's welfare in these decisions. The judge also emphasised, once again, all of the developments in law and society since the *Briody* decision, stating that 'courts have bent over backwards to recognise the relationships created by surrogacy, including foreign commercial surrogacy'.[25] Significantly, Lady Hale held that 'it is no longer contrary to public policy to award damages for the costs of a foreign commercial surrogacy', subject to a number of conditions:

> First, the proposed programme of treatments must be reasonable ... Second, it must be reasonable for the claimant to seek the foreign

21 Ibid [40].
22 Ibid [44]; see also *XX v Whittington Hospital NHS Trust* [2017] EWHC 2318 (QB).
23 *Whittington Hospital* (n 12 above) [46].
24 Ibid [48].
25 Ibid [52].

> commercial arrangements proposed rather than to make arrangements within the UK. This is unlikely to be reasonable unless the foreign country has a well-established system in which the interests of all involved, the surrogate, the commissioning parents and any resulting child, are properly safeguarded. Third, the costs involved must be reasonable.[26]

Lord Carnwath and Lord Reed dissented on the issue of recovering damages for a commercial surrogacy arrangement in a country where this is lawful. According to Lord Carnwath, while 'this case is not concerned with illegality as such, the underlying principle of coherence or consistency in the law is of broader application'.[27] Acknowledging the legal and social changes which Lady Hale discussed, Lord Carnwath concluded that '(t)here has however been no change to the critical laws affecting commercial surrogacy, which led to the refusal in 2001 of damages on that basis. Nor does the Law Commission propose any material change in that respect.'[28] Further, he concluded that:

> It is also apparent from recent studies that public attitudes remain deeply divided ... So long as that remains the state of the law on commercial surrogacy in this court, it would not in my view be consistent with legal coherence for the courts to allow damages to be awarded on a different basis.[29]

DISCUSSION

As part of the decision, Lady Hale described the current law on surrogacy as well as the plans for reform, as detailed in the Law Commissions' Consultation Paper. In a striking statement regarding domestic UK law governing surrogacy, Lady Hale noted that 'it is scarcely surprising that the claimant's clear preference is for a commercial surrogacy arrangement in California'.[30] Indeed, Lady Hale cited Sir Nelson in his judgment in this case before the High Court, where he described the appeal of a surrogacy abroad as opposed to a domestic surrogacy: 'the system is well-established, the arrangement binding and the intended parents can obtain a pre-birth order from the Californian court confirming their legal status in relation to the surrogate child'.[31] These observations by Lady Hale and by Sir Nelson summarise why intending parents may choose to engage in a commercial surrogacy abroad as opposed to a domestic, altruistic arrangement. This must

26 Ibid [53].
27 Ibid [64].
28 Ibid [67].
29 Ibid [67].
30 Ibid [22].
31 *XX v Whittington Hospital NHS Trust* (n 22 above) [31].

be considered by legislators in any reform of the law. Indeed, a recent report by the All-Party Parliamentary Group (APPG) on surrogacy recommended that '[a]n important imperative for law reform should be to create a more stable system in the UK which removes the push factors for seeking surrogacy overseas'.[32] This was also accentuated by the Special Rapporteur on Child Protection in Ireland, who recommended that regulation of surrogacy 'should incentivise reliance on domestic arrangements by adopting a more streamlined and less burdensome framework than for international arrangements'.[33]

What is also evident from this case, as well as from the reports of the UN Special Rapporteur, the Hague Conference on Private International Law and the Law Commissions' Consultation Paper, is that categorising a surrogacy arrangement as altruistic or commercial is not necessarily helpful. The UN Special Rapporteur has emphasised the importance of regulating surrogacy on both a national and international level but is critical of 'the development of organized surrogacy systems labelled "altruistic", which often involve substantial reimbursements to surrogate mothers and substantial payments to intermediaries' which 'may blur the line between commercial and altruistic surrogacy'.[34]

Jackson previously commented, for example, that 'a non-commercial surrogacy arrangement in the UK can cost as much as £35,000. If substantial sums of money are able to change hands in non-commercial surrogacy, the ban on commercial surrogacy looks rather weak and ineffective'.[35] Fenton-Glynn and Scherpe observed that 'payment for services is being hidden'.[36] Indeed, this seems to be at the crux of the issue, and it is now necessary to provide a 'better definition of what constitutes a "reasonable expense"'.[37] The APPG found that '[t]here was no real support for US-style payments to surrogates. If anything, a modest sum at most was supported.'[38]

Also, from a 'coherence' point of view, it is difficult to reconcile a prohibition in law on domestic commercial surrogacy, while in practice such commercial arrangements appear to be taking place and

32 Andrew Percy MP, *Report on Understandings of the Law and Practice of Surrogacy* (All-Party Parliamentary Group on Surrogacy2020) 'Recommendation no 8', 25.

33 C O'Mahony, *A Review of Children's Rights and Best Interests in the Context of Donor-assisted Human Reproduction and Surrogacy in Irish Law* (Government of Ireland 2020), 48.

34 A/HRC/37/60, 16, para 8.

35 Jackson (n 9 above) 206.

36 Fenton-Glynn and Scherpe (n 10 above) 4.

37 Surrogacy UK, *Surrogacy in the UK: Further Evidence for Reform* Second Report (Surrogacy UK Working Group on Surrogacy Law Reform 2018) 7.

38 Percy (n 32 above) 'Recommendation no 9', 25.

international commercial surrogacy arrangements are also consistently approved.

Finally, Lord Carnwath's description of public attitudes as remaining 'deeply divided' is noteworthy given the split within the Supreme Court itself regarding the recovery of damages to fund an international commercial surrogacy arrangement.[39] This split is indicative of the challenges faced by individual states, as well as by the Hague Conference on Private International Law, in trying to achieve consensus on a national and an international level. From the claimant's point of view, however, the majority judgment of the Supreme Court ensures that she can seek to build her family through surrogacy, despite repeated failures in the health system which deprived her of the ability to bear a child.[40]

CONCLUSIONS

The judgments delivered in this case provide an important viewpoint on international commercial surrogacy. How these judgments will shape the proposed new surrogacy law in not only the UK, but also in Ireland and internationally, remains to be seen. This article sought to highlight some important issues which arose in the judgments that must be considered by legislators for the regulation and reform of surrogacy law. Consultations with key stakeholders carried out by the APPG and the Law Commissions are hugely important to ensure that any reform in this area is informed and fit-for-purpose. As Michael Freeman commented more than two decades ago, '(w)e need to make up our minds about surrogacy'.[41]

39 For further commentary on this point, see Horsey and Powell (n 13 above).

40 It was recently reported that a man sought costs in the High Court in Ireland for surrogacy in the United States following the death of his wife from cervical cancer. This case was settled, and the terms of settlement were not disclosed. It is likely, however, that similar cases will arise again before Irish and UK courts. See Vivienne Traynor, 'Husband to use money from High Court settlement for surrogacy' (RTE, 4 March 2021).

41 M Freeman, 'Does surrogacy have a future after Brazier?' [1999] 7(1) Medical Law Review 1, 20.

Northern Ireland
Legal Quarterly
Autumn Vol. 72 No. 3 (2021) 596–604
Commentaries and Notes DOI: 10.53386/nilq.v72i3.970

Secretary of State for Justice v A Local Authority and others: disability and access to sex workers

Martin F Regan

School of Law, Queen's University Belfast

Kevin J Brown

School of Law, Queen's University Belfast
Correspondence email: m.regan@qub.ac.uk

ABSTRACT

This is a commentary on *Secretary of State for Justice v A Local Authority and others*, where the decision of the Court of Protection has been overturned by the Court of Appeal. The judgment has implications for (i) the article 8 and article 14 rights of those who lack capacity to arrange lawful sexual services; (ii) the criminal liability of their carers who are enlisted to assist with such arrangements; and, potentially, (iii) the ban on payment for sexual services in Northern Ireland.

Keywords: sex worker; mental capacity; care worker; human rights; statutory interpretation; Sexual Offences Act 2003; Sexual Offences (Northern Ireland) Order 2008.

INTRODUCTION

In *A Local Authority v C and others*,[1] the Court of Protection in England and Wales held that an individual who lacks capacity to organise services from a sex worker could, in theory, enlist the help of their care workers to make the necessary arrangements, without the latter facing criminal liability under section 39 of the Sexual Offences Act 2003 (the offence of care workers causing or inciting sexual activity). The Court of Appeal rejected the lower court's interpretation of section 39 in the context of the circumstances envisaged in the case.[2] The judgment has implications for article 8 (right to privacy) and article 14 (prohibition of discrimination)[3] for those lacking capacity, in matters of arranging sexual services that are lawfully available to those who do not require

1 [2021] EWCOP 25.
2 [2021] EWCA Civ 1527.
3 European Convention for the Protection of Human Rights and Fundamental Freedoms 1950.

such assistance. The case also has potentially wider implications on the general prohibition in Northern Ireland on paying for sexual services.

BACKGROUND

The case before the Court related to C, a young man who wished to engage in sexual activity with a sex worker. He had been diagnosed with a genetic disorder, Klinefelter syndrome (XXY syndrome), which manifested by the age of two, in the form of developmental delay and communication difficulties. At the age of four, he was diagnosed with autistic spectrum disorder. Subsequently, C's behaviour became sometimes challenging and aggressive, and he was required to move out of the family home. The result is that C needs significant assistance with independent living and the support requires the deprivation of his liberty. This has been authorised by the Court of Protection since 2017, prior to which (from 2014–2017), C was detained in hospital under the Mental Health Act 1983. Due to progress in C's treatment, it was possible to discharge him to his current home, which is a house suitable for three occupants and their carers. An agency provides a support package to meet C's needs. In August 2018, C told AB, his litigation friend, that he wanted to be able to have sex, but did not think he had much prospect of finding a girlfriend. He wanted to know if his carers would assist him in making contact with a sex worker. It was agreed by all parties that C had the capacity to engage in sexual activity and to decide to do so with a sex worker, but lacked the ability to contact a sex worker for himself. The local authority commenced proceedings to address the lawfulness of C's carers assisting him in accessing sexual services.[4]

The issues before the court were: 'whether a care plan to facilitate C's contact with a sex worker could be implemented without the commission of an offence under the Sexual Offences Act 2003'; 'if not, whether the Sexual Offences Act 2003 can be read compatibly with the European Convention on Human Rights, or whether the court should make a declaration of incompatibility'; and 'if a care plan facilitating such contact is lawful, whether such a plan would be in C's best interests'.[5]

COURT OF PROTECTION DECISION

Hayden J considered the legal framework, specifically section 39 on care workers inciting sexual activity and the definition of the former.[6] It was undisputed that C had a mental disorder and that those who

4 [2021] EWCOP 25, [1]–[5].
5 Ibid [6].
6 Sexual Offences Act 2003, s 42(4).

would potentially assist C were care workers. Offences relating to soliciting[7] and paying for sex from someone subject to force[8] were not considered relevant as it was not intended to procure sexual services in these circumstances. There followed an examination of the section 39 offence and the mischief it was intended to address, noting that paying for sex itself is not an offence.[9] Counsel for C argued that the sort of assistance envisaged in making the practical arrangements for an encounter with a sex worker fell outside the scope of section 39 and would not incur criminal liability. By contrast, the Secretary of State argued that such an interpretation would amount to a change in the law and undermine parliamentary sovereignty.[10]

There was expert testimony from Professor Claire De Than, a legal academic, who is involved with The Outsiders Trust charity incorporating the TLC Trust, which provides support for individuals with disabilities, in the matter of sexual and intimate services. This highlighted the therapeutic value of facilitating a transition towards personal sexual relationships,[11] whilst detailing the rules and policies in place to protect both service providers and users.[12]

It was held that: facilitation of C's contact with a sex worker would not constitute an offence under section 39 of the Sexual Offences Act 2003, when interpreted as intended by Parliament, giving the words their natural and obvious meaning; it was not necessary to invoke section 3 of the Human Rights Act 1998 in order to construe section 39 as being compatible with the Convention because the natural meaning of the words, and the purpose of the statute as a whole, would not infringe C's article 8 rights; whether the proposed plan was in C's best interests would be considered at a later date, following a risk assessment.

The court rejected assertions that the issue had been determined previously,[13] in *Lincolnshire County Council v AB*,[14] but acknowledged that a clarification of conflicting interpretations of section 39 was merited, lest the court's interpretation conferred a seal of approval on prostitution, contrary to public policy. Due to the potential conflict between general policy considerations in relation to prostitution and the proper interpretation of section 39 in the instant case, the court granted the Secretary of State permission to appeal. Hayden J emphasised that this was specifically not on the basis of there being

---

7 Ibid s 51A.
8 Ibid s 53A.
9 [2021] EWCOP 25, [16]–[22].
10 Ibid [37].
11 Ibid [27].
12 Ibid [30]–[36].
13 Ibid [23].
14 [2019] EWCOP 43.

any 'real prospect of success',[15] but solely because the interpretation of 'intentionally causes or incites'[16] fell within the category of cases where there was 'some other compelling reason for the appeal to be heard'.[17] Hayden J's pronouncement on the prospects of the appeal proved to be a hostage to fortune.

COURT OF APPEAL DECISION AND COMMENTARY

In delivering his judgment in a unanimous decision, Lord Burnett CJ highlighted the essence of the reasoning at first instance, namely that the words in section 39 were aimed at those in a position of authority and trust, who sought to undermine the sexual autonomy of those with a mental disorder.[18] Concerns were expressed regarding the hypothetical nature of the situation, since no 'order' was being appealed and that, whilst declarations on the legality of a care plan are permitted,[19] this should be confined to exceptional circumstances, where potential transgressions of the criminal law are concerned.[20] Focusing on section 39, Lord Burnett CJ was clear that the arrangements proposed would constitute legal causation of the sexual activity, rather than merely creating the circumstances in which this could occur. The latter would be characterised by arranging contact between an individual and their spouse/partner, during which sexual activity might more 'naturally' take place.[21] In this sense, the court was of the view that Hayden J had erred in his interpretation of 'causes or incites', favouring instead the decision in *Lincolnshire County Council v AB*.[22]

On the European Convention on Human Rights (ECHR) aspects, the court was equally unambiguous in determining that there was no positive obligation on the state, under article 8, to permit care workers to arrange for sexual contact with prostitutes,[23] and that, even if there were any interference with individual rights, this would be justified under article 8.2.[24] Lord Burnett CJ was unsympathetic to article 14 arguments, stating that the discriminatory effect of section 39 is justified on the grounds that Parliament's considered intention was to provide a 'cloak of protection' for vulnerable individuals.[25]

15 CPR 52.6 (1)(a).
16 Sexual Offences Act 2003, s 39(1)(a).
17 CPR 52.6 (1)(b).
18 [2021] EWCA Civ 1527, [23].
19 Mental Capacity Act 2005, s 15.
20 [2021] EWCA Civ 1527, [30].
21 Ibid [49].
22 See n 14 above.
23 [2021] EWCA Civ 1527, [53].
24 Ibid [60].
25 Ibid [64].

The Court of Appeal has taken a cautious approach in this case, in furtherance of the uncontentious and, indeed, laudable goal of protecting the vulnerable. It is argued, however, that there are problems with the approach taken, not least in the realm of personal autonomy and the role of the 2005 Act[26] in promoting such. The judgment also appears to allude obliquely to the morality of specific types of sexual relationship, in the context of differentiating between causing versus creating the circumstances for an encounter, for example, King LJ refers to 'less extreme and benign situations' and to the circumstances being different for 'a long married couple'.[27] This hints at a certain unacceptability of a relationship that is purely sexual (including in the transactional sense, as in C's case).

Interestingly, the court did not permit the Secretary of State to amend the grounds of appeal to include the stipulation that for the courts 'to sanction the use of a sex worker is contrary to public policy'.[28] Furthermore, Baker LJ, whilst concurring with the Lord Chief Justice and Lady Justice on the issue of the circumstances under which facilitation of a sexual encounter could be permissible, emphasised that the court was concerned only with the judge's decision in C's case and that a declaration under section 15 in relation to a care plan will turn ultimately on the specific, detailed facts.

The Court of Protection had attempted to reinforce and assert the autonomy of individuals who, whilst lacking capacity to make the practical arrangements necessary to receive sexual services, face no such impediment in expressing their wish to receive those services. To deny such individuals access to lawful sexual services is a form of discrimination based on disability and a violation of their article 8 and article 14 rights. The Court of Appeal's view is that this is either not the case or justified, respectively.

If the rights of C are to be upheld, this inevitably leads to a determination as to the potential liability of C's care workers in facilitating those rights. The court at first instance was unambiguously of the view that the intention of Parliament is clear when the 2003 Act is read in its entirety and the words given their literal meaning.[29] Furthermore, the 'mischief'[30] that Parliament intended to suppress focuses on the sexual exploitation of the vulnerable, including situations where there is a breach of a relationship of care, as detailed

26 Mental Capacity Act 2005.
27 [2021] EWCA Civ 1527, [71], [75] (Baker LJ).
28 Ibid [5].
29 [2021] EWCOP 25, [44]; *Fisher v Bell* [1960] 3 All ER 731.
30 *Pepper v Hart* [1993] AC 593.

in the White Paper[31] preceding the introduction of the Sexual Offences Bill in November 2002.

Notwithstanding the Court of Appeal judgment, it is asserted that, whichever rule of statutory interpretation is preferable, the words 'intentionally causes or incites' mean just that. In this regard, it would appear to be C, rather than his care workers, who is doing the 'causing' and 'inciting', therefore the actions of the latter in facilitating C's wishes should fall outside the scope of the legislation. The situation would be entirely different had the proposition been instigated by the care worker, in which case they would clearly fall foul of section 39; that is not what occurred in this case (indeed, nothing occurred). Taking the 2003 Act as a whole, it is clear that the words at issue refer to activity of an exploitative character. It also seems clear that frustrating the desire of an individual to engage in sexual activity was not Parliament's intention when enacting the legislation. Rather, the legislation, and the White Paper that preceded it, represent a concerted effort to tackle the sexual exploitation of the vulnerable, which is rightly viewed as an area for legislative action.

The hypothetical nature of C's case (since no care plan had yet been put in place by the court) probably did not help at appeal. Nevertheless, the judgment raises questions about what sort of assistance would constitute legal causation. In the aftermath of the judgment, commentary from Junior Counsel for C seems apposite, particularly in relation to what would constitute 'causing', in practical terms, for example: setting aside money so that an individual can access a sex worker; helping an individual into bed in advance of the arrival of a sexual partner; making best interests' decisions whereby a prospective relationship is anticipated to be sexual?[32]

WIDER IMPLICATIONS FOR THE LAW IN NORTHERN IRELAND

At first sight, the judgment would appear to close down further debate in the Northern Ireland context. If a similar request were to be made by a person who resides in that jurisdiction, the courts would determine that paying for in-person sexual services is a criminal offence under article 64A of the Sexual Offences (Northern Ireland) Order 2008, whatever the circumstances. This means that the equivalent of C would face potential criminal liability, as well as the carers. The court

31 Home Office, *Protecting the Public: Strengthening Protection Against Sex Offenders and Reforming the Law on Sexual Offences* (Cm 5668, 2002).

32 Ben McCormack, 'Re C – the Court of Appeal's view' (Garden Court North Chambers, 22 October 2021).

in *A Local Authority v C and others* stated that its decision was based on the fact that paying for sex is not *per se* illegal in England and Wales; however, it is not too much of a stretch of the jurisprudential imagination to envisage that, prior to the Court of Appeal judgment, the case might have inspired a challenge to the broader prohibition found in Northern Ireland. Such a case would likely involve recognising that some individuals with disabilities may decide that the only realistic opportunity for having a sexual relationship with another adult is to pay for it and that the law should not prohibit them from doing so where someone is willing to provide such a service. The Court of Appeal judgment may have, for the time being, neutralised this question from a judicial perspective, however, the issue is unlikely to evaporate and, as outlined above, the decision raises further questions regarding what would be permissible in practice.

Studies from several jurisdictions have shown that demand for access to sexual services amongst disabled people exists.[33] In some countries where sex work is legal, a distinct profession is developing called sexual assistants. These are men or women 'of any sexual orientation who, after professional training, can engage in sexual activity with persons with any type of disability'.[34] In the UK, the TLC Trust, provides online listings of sex workers in Great Britain who provide services to those with disabilities. The mission of the charity is

> that disabled people can use sexual and intimate services to help them learn about physical pleasure and may enable them to move forward towards personal sexual relationships. Where this is not possible, we would like to ensure that all disabled people have access to sexual, sensual and intimate experiences.[35]

The right to form relationships including consensual sexual relations with other human beings is recognised within the concept of private life under article 8.[36] The broad prohibition on paying for sex found in Northern Ireland is arguably a breach of that right as it criminalises all forms of sex work, including the work of sexual assistants. The disproportionate impact on those with disabilities also potentially brings into play article 14, which requires that the rights set out in the Convention are protected and applied without discrimination. Such reasoning could lead to the conclusion that, in legislating to prohibit

33 G R Gammino, E Faccio and S Cipolletta, 'Sexual assistance in Italy: an explorative study on the opinions of people with disabilities and would-be assistants' (2016) 34(2) Sexuality and Disability 157; M Girard, M T M Sastre and E Mullet, 'Mapping French people's views regarding sexual assistance to people with physical disabilities' (2019) 37(1) Sexuality and Disability 109.

34 Gammino et al (n 33 above) 157.

35 The TLC Trust, 'What is TLC?'.

36 *Pretty v UK* (App no 2346/02) ECHR 2002, para 61.

the payment for sexual services in all circumstances, the Northern Ireland Assembly acted in contravention of the ECHR and in doing so acted *ultra vires*, therefore leading to the striking down of article 64A. A previous challenge by a sex worker against the legislation in Northern Ireland was given permission by the High Court to proceed to a judicial review, but the proceedings were dropped when the applicant died prior to the full hearing.[37] Laura Lee, the sex worker in question, had spoken about the importance she attached to the role of providing sexual fulfilment to clients with disabilities.[38]

Any decision by the courts to find a right to sexual services would be controversial. Some disabled people might view such as a decision as promoting the stigmatising myth that the only sexual fulfilment that those with disabilities can have is by paying for it. However, accepting the reality that a significant proportion of disabled people face sexual marginalisation is not to argue that this is true of all disabled people.[39] Meanwhile, those who favour the criminalisation of sex work would also naturally find any recognition of a right to access such services deeply problematic. Julie Bindel, academic and commentator, whilst agreeing that the first instance decision may have acted as a springboard to the recognition of a right to access sexual services for those with disabilities, warns that such a path would be a dangerous one to go down.[40] She argues that it risks 'disabled people being held up as a handy smokescreen for pimps and exploiters' whereupon such jurisprudence would ultimately lead to recognition of a general right for all adults to access sexual services.[41] In light of the Court of Appeal's decision, these concerns may now be allayed.

Opponents of recognition of a right to sexual services for those with disabilities will rely on article 8 being a qualified rather than an absolute right. If a challenge as suggested above was brought before the courts in Northern Ireland, the parties defending article 64A would presumably argue that a broad prohibition on the paying for sexual services is in accordance with the law, furthers the legitimate aims of the prevention of crime, the protection of health or morals, or the protection of the rights and freedoms of others and is necessary and proportionate in achieving that aim. A counter to such arguments is that

37 H McDonald, 'Irish sex worker and campaigner for rights of prostitutes dies, aged 39' *The Guardian* (London, 9 February 2018).

38 M McGrath, '"We bring happiness into their lives" – meet the sex workers providing services for clients with disabilities' *Irish Independent* (Dublin, 12 September 2016).

39 S Esmail, K Darry, A Walter and H Knupp, 'Attitudes and perceptions towards disability and sexuality' (2010) 32(14) Disability and Rehabilitation 1148.

40 J Bindel, 'Disabled men don't have a "right" to buy sex' *The Spectator* (30 April 2021).

41 Ibid.

the broadness of the Northern Ireland prohibition is disproportionate to the legislation's stated aim of protecting vulnerable individuals from sexual exploitation.[42] Indeed, the Northern Ireland legislation has been criticised as counterproductive in that sex work continues, but with sex workers at greater risk of harm by forcing them and those who pay for their services underground and out of sight from the protection of the authorities.[43] Therefore, a more proportionate and effective approach to achieving the aim of protecting the vulnerable would be to criminalise those who pay for the services of controlled or coerced sex workers as is the case in England and Wales.[44] Such a decision, whilst controversial, would arguably provide a better balance of the competing interests of the need to respect sexual autonomy and protection from sexual exploitation.

The Court of Appeal decision, whilst unambiguous on the meaning of 'causing' in section 39, has prompted further questions about determining the source of this in cases such as C's; the issues remain unresolved, in terms of certainty around the potential criminal liability of care workers.

42 G Ellison, 'Criminalizing the payment for sex in Northern Ireland: sketching the contours of a moral panic' (2017) 57(1) British Journal of Criminology 194.

43 G. Ellison, C Ní Dhónaill and E Early, 'A Review of the criminalisation of paying for sexual services in Northern Ireland' (Department of Justice 2019).

44 Sexual Offences Act 2003, s 53A.

Northern Ireland
Legal Quarterly
Autumn Vol. 72 No. 3 (2021): 605–611
Commentaries and Notes DOI: 10.53386/nilq.v72i3.971

Confiscation orders, confusion and a lament for the past: the case of Bernadette Hilton: *R v Hilton* [2020] UKSC 29

John E Stannard

School of Law, Queen's University Belfast
Correspondence email: j.stannard@qub.ac.uk

The last few years have witnessed some memorable cases in the Supreme Court, dealing with such momentous issues as Brexit,[1] the right of the Prime Minister to prorogue Parliament,[2] and the detention of Gerry Adams under the 1972 internment legislation.[3] However, even the most trivial and seemingly humdrum of cases can end up in that august tribunal. When Bernadette Hilton was convicted of benefit fraud in September 2015, she cannot have imagined that her case would end up four years later in the highest court in the land.[4] Yet, though the case might have appeared routine in nature, it raises a number of fundamental and wide-ranging issues with regard to the making of confiscation orders, the sentencing regime generally and the relationship between the legal academy and the professions.

The facts of the case are relatively simple, but the legal issues to which it gave rise were anything but.[5] The defendant was convicted before the magistrates' court in Belfast of making false statements in order to obtain income support, contrary to section 105A of the Social Security Administration (NI) Act 1992. She was then committed to the Crown Court with a view to the making of a confiscation order under section 156 of the Proceeds of Crime Act 2002 (POCA), which corresponds to section 6 of that Act in relation to England and Wales.[6] Though the provisions regarding this are extremely complex,[7] the basic rule is that the court has to calculate the extent of the benefit

1 *R (on the application of Miller and another) v Secretary of State for Exiting the European Union* [2017] UKSC 5.
2 *R (on the application of Miller) v The Prime Minister* [2019] UKSC 41.
3 *R v Adams* [2020] UKSC 19.
4 *R v Hilton* [2020] UKSC 29, [2020] 1 WLR 2945.
5 The facts are taken from paras [1]–[3] of the judgment.
6 Pt 1 of the Act deals with England and Wales and pt IV of this deals with Northern Ireland. Since the issues in the case apply equally to England and Wales, the corresponding English section will also be given in the relevant footnote.
7 See, generally, Archbold Criminal Pleading Evidence and Practice (2021 edn) chapter 5B.

gained by the defendant from the relevant criminal conduct,[8] and then to make a confiscation order in respect of that sum.[9] However, in many cases the defendant may not have the means to pay the full sum,[10] and here the court must make an order up to the limit of his or her available assets.[11] And this is precisely what happened in the case of Ms Hilton. The extent of the benefit was assessed at £16,517.59, but the defendant did not have the resources to pay this in full. So the court went on to calculate her available assets and came up with a figure of £10,263.50, based on the value of a house held by the defendant jointly with her former partner, less a sum still owing in respect of an outstanding mortgage. So far, so good.

However, this is where things began to get complicated. The defendant appealed against the making of the order,[12] and while this was being prepared someone drew her attention (or rather that of her counsel) to section 160A[13] of the 2002 Act, which had been inserted by virtue of section 24 of the Serious Crime Act 2015 and had come into force at the beginning of June that year. The key provisions of section 160A were (and are) as follows:

> (1) Where it appears to a court making a confiscation order that—
>
> (a) there is property held by the defendant that is likely to be realised or otherwise used to satisfy the order, and
>
> (b) a person other than the defendant holds, or may hold, an interest in the property,
>
> the court may, if it thinks it appropriate to do so, determine the extent (at the time the confiscation order is made) of the defendant's interest in the property.
>
> (2) The court must not exercise the power conferred by subsection (1) unless it gives to anyone who the court thinks is or may be a person holding an interest in the property a reasonable opportunity to make representations to it.

---

8 POCA, s 156(4) (s 6(4) for England and Wales). Normally, the court will look in this connection at the conduct relating to the offence of which the defendant stands convicted; this is called his or her 'particular' criminal conduct (s 156(4)(c) (s 6(4)(c)). But in cases where the defendant is found to have a 'criminal lifestyle', the court can also take into consideration his or her 'general' criminal conduct, which has a much wider focus (s 156(4)(b) (s 6(4)(b)). However, there was no question of a criminal lifestyle in the present case.

9 POCA, s 157(1) (s 7(1) in England and Wales).

10 The burden of proof here is on the defendant: POCA, s 157(2) (s 7(2) in England and Wales).

11 This is called 'the available amount': POCA, s 157(2)(a) (s 7(2)(a) in England and Wales). S 159(1) (s 9(1)) sets out the formula by which this is to be calculated.

12 *R v Hilton* (n 4 above) para [3].

13 Corresponding to s 10A for England and Wales.

This provision had not been cited to the Crown Court at the time the order was made,[14] but the defendant argued that it was fatal to the validity of the order. In particular, she highlighted section 160A(2) in this connection.[15] As we have seen, the decision of the Crown Court as to the available sum was based on the value of a house held by the defendant jointly with her former partner, which was moreover still subject to an outstanding mortgage. Yet, neither the partner nor the mortgagee had even been aware of the proceedings, still less had they been given any opportunity to make representations at the time when the confiscation order was made.

The defendant's contentions in this respect were upheld by a unanimous Court of Appeal.[16] The Crown Court judge, they concluded, had clearly made a determination under section 160A(1), but had overlooked section 160A(2).[17] In the words of Deeny J:[18]

> [T]he language used by Parliament would suggest that it was intended that this be a mandatory provision and the court having exercised its power under sub-section 1 ought to have done that. In any event the provision is a sensible one in case there had been some development since the title to the property had been commenced which was not reflected on the title to the property and by which one of the other persons with an interest the property, in this case the estranged husband and the lender, might be able to persuade the court that this appellant did not have a 50% interest in the property but conceivably a larger or a smaller interest either of which would affect the order to be made by the court. The omission to do that we consider is fatal to the decision of the judge.

Subsequently, a further appeal was brought by the Director of Public Prosecutions, the following point of law being certified by the Court of Appeal:

> 1. Where property is held by the defendant and another person, in what circumstances is the court making a confiscation order required by section 160A of the Proceeds of Crime Act 2002, in determining the available amount, to give that other person reasonable opportunity to make representations to it at the time the order is made?
>
> 2. If section 160A does so require, does a failure to give that other such an opportunity render the confiscation order invalid?

Now, one would have thought at first sight that the answer to this question was obvious – indeed so obvious that it hardly merited the attention of the Court of Appeal, let alone the Supreme Court. After all,

---

14 *R v Hilton* (n 4 above) para [6].
15 Corresponding to s 10A(2) for England and Wales.
16 *R v Hilton* (n 4 above) paras [5]–[6].
17 Ss 10A(1) and 10A(2) for England and Wales.
18 *R v Hilton* [2017] NICA 73 para [7].

is not one of the most fundamental principles of natural justice *audi alteram partem*, one implication of which is that no person's rights – whether or not they are actually party to the case – should be affected without giving them an opportunity to be heard? Moreover, section 160A(3)[19] goes on to say that any determination made under the section shall be 'conclusive' in relation to any question as to the extent of the defendant's interest in the property that arises in connection with the realisation of the property, or the transfer of an interest in the property with a view to satisfying the confiscation order, or any action or proceedings taken for the purposes of any such realisation or transfer! Yet, the appeal was allowed unanimously by the Supreme Court in the present case, and the original confiscation order upheld. How can this be?

The answer is that the Crown Court had been quite right to ignore section 160A here, and that both the defendant and the Court of Appeal had misunderstood that provision. To understand this involves a fairly detailed analysis of the legislative history of the provision in question. This was duly undertaken by the late Lord Kerr of Tonaghmore, who handed down the definitive judgment in the case.

The key to the whole matter, as Lord Kerr pointed out, was the distinction between two stages of the confiscation regime under the 2002 Act, one being the *making* of the order and the other its *enforcement*. Prior to the introduction of section 160A in 2015, the picture was clear. The making of the order was governed by sections 156–163[20] and, as we have seen, involved the court in calculating the relevant benefit, and then making an order that the defendant pay that sum, or the available amount if less. This was intended, as Lord Kerr pointed out, to be a fairly straightforward if not automatic process. In particular, there was no question of third parties having to be consulted. Why was this? The answer is because the making of the order did not affect such parties in any way. All it did was to create a statutory debt payable by the defendant to the court. In the words of *Millington and Sutherland Williams*, a leading practitioner text,[21] a confiscation order was no more than 'an in personam order against the convicted defendant'. It was not 'an in rem order against specific items of property'.

In most cases, no doubt, the intention was that the order would duly be paid by the defendant and that no more would be heard of it. However, if this were not done, then the 2002 Act provided a

19 S 10A(3) for England and Wales.

20 Ss 6–13 for England and Wales.

21 *Millington and Sutherland Williams on the Proceeds of Crime* 8th edn (Oxford University Press 2018) 16.53.

machinery for enforcement of the order. In particular, section 198[22] allowed for the appointment of a receiver to deal with and, if necessary, realise the defendant's assets. Unlike the making of the original order, this, of course, might very well affect the interests of third parties, and section 199(8)[23] of the Act catered for this by providing that the court should not confer or exercise these powers without giving persons holding interests in the property a reasonable opportunity to make representations to it.

So where did section 160A[24] come into the picture? This, as we have seen, was not in the original Act at all, but was introduced some 13 years later by the Serious Crime Act 2015. The reason for its introduction, as Lord Kerr explained, was to provide an abbreviated procedure combining the confiscation and enforcement stages in simple cases where there could be no sensible debate about how the confiscation order should be enforced, and where there was therefore no point in going through the whole gamut of the two-stage process. But in other cases, where the issues were clearly not so simple, both stages would continue to apply, and here any representations made by third parties would have to wait until the second stage, as in times past.

So, how did this work out in terms of the language used by section 160A? The answer was that the requirements in section 160A(2) only came into play in cases where the court exercised 'the power conferred by subsection (1)'. But in the present case the Crown Court had done no such thing. Yes, it had worked out the 'available amount' using the formula under section 159,[25] but that was not the same as making a determination under section 160A. Since the court had never exercised the power conferred by section 160A(1), section 160A(2) had no application to the case.

In sum, the position with regard to confiscation orders, as envisaged by the Supreme Court, seems to be as follows. In most cases, section 160A[26] will have no application, and the normal two-stage process will continue to be followed. In these cases the making of the order under section 156[27] will only take effect *in personam*, and therefore the rights of third parties need not be considered unless and until proceedings have to be taken for the order to be enforced under section 199.[28] However, some cases may be sufficiently straightforward for the court to apply the streamlined procedure under section 160A; where this is

22 S 50 for England and Wales.
23 S 51(8) for England and Wales.
24 S 10A for England and Wales.
25 S 9 for England and Wales.
26 S 10A for England and Wales.
27 S 6 for England and Wales.
28 S 51 for England and Wales.

done, the order will take effect not only *in personam* but *in rem*, and therefore others with an interest in the property will have to come on board at the outset.

In so far as the Supreme Court has provided clarity on this issue, the decision is to be welcomed. However, as indicated above, there are broader issues at stake. It might seem odd that the key provision here – section 160A[29] of the 2002 Act – was totally overlooked by the court at first instance and was then misapplied by the Court of Appeal. However, it has been notoriously difficult for the courts and the professions to keep up with legislative changes in the area of sentencing. Back in 2015, Andrew Ashworth referred to the 'complexity and relentless frequency' of much recent sentencing legislation[30] and followed David Thomas in highlighting 'the omissions and confusion resulting from late amendments, defective drafting, legislation by incorporation, staggered commencement dates and ill-conceived transitional provisions'.[31] What made this even worse in the case of sentencing was the need for the courts, in the light of article 7 of the European Convention on Human Rights, to keep in mind not only the current law but the law that was in place at the time when the offence was committed. No wonder that they have sometimes got the law wrong, though hopefully the new Sentencing Code will improve matters from now on, at least as far as England and Wales is concerned.[32]

This, of course, is where the academic profession comes in. Had section 160A been introduced prior to 2012, it would most certainly have been picked up and explained by the *Bulletin of Northern Ireland Law*, a digest of current legal developments in Northern Ireland which in its own words allowed for 'easy browsing and searching and offered links to the full text of selected legislation and written judgments and other reference material'.[33] Alas, this excellent resource is no more, having ceased publication nine years ago with the demise of the 'Servicing the Legal System' (SLS) project set up by Queen's and the professions in 1981. This was a retrograde step, to say the least. One of the greatest strengths of the Queen's Law School in the past lay in its close relationship with the professions and in the academic support provided through SLS and other channels. Of course, one cannot say

29 S 10A for England and Wales.

30 Andrew Ashworth, *Sentencing and Criminal Justice* 6th edn (Cambridge University Press 2016) 1.5.1.

31 Ibid; David A Thomas, 'Sentencing legislation – the case for consolidation' [1997] Criminal Law Review 406.

32 See now the provisions of the Sentencing Act 2020. These, however, do not contain the provisions relating to confiscation orders, which continue to be governed by the POCA.

33 Bulletin of Northern Ireland Law, University of Ulster Library.

that the problems arising in the case under discussion would have been prevented had the *Bulletin* and SLS still been in existence, but certainly there would have been more chance of the crucial point being picked up.

There is no point in regretting the good old days, but certainly there is an argument for more co-operation between the Law School and the professions, and the case of *Hilton* provides good support for it.

Northern Ireland
Legal Quarterly
Autumn Vol. 72 No. 3 (2021) 612–621
Commentaries and Notes DOI: 10.53386/nilq.v72i3.972

Being transgender: human variation or disorder?

Alexandra Maxwell

School of Law, Queen's University Belfast
Correspondence email: amaxwell14@qub.ac.uk

INTRODUCTION

Eekelaar describes the area of law, in which people transition to a different gender, as an 'evolving understanding of reality',[1] a helpful lens through which to view the present case. The judgment in the case of *Re JR111*[2] asserts that being transgender is neither a mental illness nor a disorder and, as such, addressed why a diagnosis of gender dysphoria[3] remains a requirement to secure a gender recognition certificate (GRC) under the Gender Recognition Act 2004 (the 2004 Act). Undeniably, safeguards are important because changing gender significantly alters one's legal status,[4] granting a successful GRC applicant new legal documentation such as a birth certificate reflecting the acquired legal gender.[5] The requirement for a gender identity disorder diagnosis 'irrationally requires transgender people to say that their understanding of their gender is caused by a mental disorder rather than a normal function of human variation',[6] stigmatising transgender people.[7]

FACTS

The applicant, who had lived in Northern Ireland as a woman since 1999, sought a GRC for legal recognition of her acquired gender. Under the 2004 Act, applicants must provide a report from a registered medical practitioner or registered psychologist practising in the field of gender dysphoria, including details of the diagnosis of the applicant's gender dysphoria. The applicant stated that this requirement, in legal

1 J Eekelaar, 'The law, gender and truth' (2020) 20(4) Human Rights Law Review 797, 808.
2 [2021] NIQB 48.
3 Defined in the 2004 Act as 'the disorder variously referred to as gender dysphoria, gender identity disorder and transsexualism'.
4 *JR111, Re Application for Judicial Review* [2021] NIQB 48, [31].
5 Ibid [32].
6 Ibid [16].
7 Ibid.

and medical terms, equated the transition from one gender to another with a recognised mental disorder, gender dysphoria, and made her feel she was 'pathological and disordered'.[8] The applicant, anonymised by court order, initially brought a wide application for judicial review against the Department of Health in Northern Ireland, but early on in the proceedings an order was made substituting the Government Equalities Office (GEO), which was responsible for the 2004 Act.[9] The Northern Ireland High Court chose to deal first with her claim that the requirements of the 2004 Act breached her right to private life under articles 8 and 14 of the European Convention on Human Rights (ECHR) – the right to respect for private and family life and the right to protection from discrimination, respectively. The High Court ruled that the requirement for an applicant to prove they have had a mental 'disorder' was incompatible with article 8, although the general requirement for a diagnosis in support of an application for a GRC was within Parliament's discretion.[10]

ISSUES

Scoffield J, delivering the judgment, addressed the following two Convention-compatibility issues; firstly, the requirement to provide a medical diagnosis and, secondly, that diagnosis being of gender dysphoria.[11] The impugned provisions of the 2004 Act were challenged as breaching the applicant's rights under articles 8 and/or 14 of the ECHR.[12]

JUDGMENT ANALYSIS

Outlining the applicant's submissions

The applicant submitted that the requirement for a gender dysphoria diagnosis was unnecessary as the remaining criteria in section 2(1) of the 2004 Act – living as one's acquired gender for two years and making a statutory declaration – amply demonstrate that a person has taken decisive steps to live fully and permanently in their acquired gender.[13] Scoffield J rejected this argument as he viewed the process through a

8 Ibid [21].
9 Ibid [17].
10 K Flood, 'Northern Ireland High Court: requirement to show medical "disorder" for gender recognition certification held incompatible with ECHR' (*Scottish Legal News*, 21 May 2021)
11 *JR111* (n 4 above) [17].
12 Ibid [19].
13 Ibid [120].

wider lens, finding several other criteria to be of equal importance.[14] The applicant particularly objected to what she maintained was the outdated and derogatory requirement of a diagnosis expressly defined to be a disorder.[15] Applicants comfortable with their transgenderism would be presented with the dilemma of either lying to obtain a diagnosis or not meeting the diagnostic criteria for gender dysphoria. Moreover, since an applicant for a GRC need only show that they, at some point, *had* gender dysphoria, the requirement lacked immediate relevance to the consideration of the applicant's circumstances at the time of the application.[16] The applicant emphasised inconsistency in the government's position, having expressed repeatedly that being transgender does not equate to being mentally ill, yet jettisoning any effort at reform to enact such sentiments.[17]

Outlining the respondent's submissions

The respondent submitted that the Joint Committee on Human Rights neither recommended the term 'gender dysphoria' cease being used, nor ruled out use of a medical element to issue a GRC. The respondent justified maintaining the requirement for a diagnosis of gender dysphoria as providing certainty and protecting the rights of others, highlighting concern over giving 'legal recognition to lifestyle changes'.[18] The criteria operate as a barrier to applicants making precipitous applications for a GRC and against 'cheating' the process,[19] as more leniency might create additional scope for abuse, particularly against vulnerable women. Although Scoffield J emphasised that in the potential, but *rare*, cases where this is done nefariously, the correct response should be to deal with the perpetrator.[20] Also, Sharpe contends that issues concerning the process of transitioning and access to single-sex spaces are unconnected, favouring the 'de-pathologisation' of legal recognition, denouncing 'bogeyman' arguments as fearmongering directed against a minority.[21] However, Nicol advocates, particularly in reference to women's single-sex spaces, that it draws in competing human rights, particularly articles 2 and 3, the right to life and

14 Ibid [134].
15 Ibid [121].
16 Ibid [122].
17 Ibid [124].
18 Ibid [126].
19 Ibid [127].
20 Ibid [130].
21 A Sharpe, 'Will gender self-declaration undermine women's rights and lead to an increase in harms?' (2020) 83(3) Modern Law Review 539, 541.

freedom from inhuman and degrading treatment.[22] He asserts that more evidence is required that harm would not be caused and denotes opposing arguments as 'theoretical and illusory', although evidence on either position is scarce.

The Government's considerations for reform of the 2004 Act

The court noted recommendations for reform of the process established in the 2004 Act which emerged in the 2016 House of Commons Women and Equalities Committee Report,[23] supporting gender self-identification and noting other countries' use of 'more enlightened' models.[24] The Committee criticised the 2004 Act's 'medical approach' and 'pathologisation' (treating transgender identities as a disease or disorder) for causing significant offence and distress for some transgender people.[25] It drew parallels to homosexuality, which was similarly classified by the World Health Organization (WHO) International Classification of Diseases (ICD) as a mental disease until 1992.[26] The ICD has now been revised regarding both homosexuality and transgenderism. The ICD-10 replaced categories of 'gender dysphoria' with 'gender incongruence' and moved the categories from the 'Mental and behavioural disorders' chapter into a new chapter entitled, 'Conditions related to sexual health'.[27] Gender incongruence is defined as being 'characterised by a marked and persistent incongruence between an individual's experienced gender and the assigned sex, which often leads to a desire to "transition", in order to live and be accepted as a person of the experienced gender'.[28] The fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5), published by the American Psychiatric Association in 2013,[29] took a similar approach, indicating international support for this change in terminology. Scoffield J utilised this in rejecting the respondent's argument that clinicians and other relevant practitioners would be unable to adapt to a terminology change.

Addressing stigmatisation, in 2020 the LGBT Health Adviser recommended that the issue of stigma be contextualised within the overall process, not only one criterion. He considered the diagnosis

22 D Nicol, 'Are trans rights human rights? The case of gender self-ID' (2021) Public Law 480, 482.

23 House of Commons Women and Equalities Committee, *Transgender Equality* (Report, HC 390, 2016).

24 *JR111* (n 4 above) [36].

25 Ibid [37].

26 Ibid [38].

27 Ibid [42].

28 Ibid [43].

29 Ibid [44].

categorisation (whether under mental or sexual health) irrelevant, as it remained a diagnosis,[30] diminishing the impact both individually and socially of such categorisation. The Department of Health and Social Care concurred, further stating that the department 'don't believe there is any stigma attached' to a diagnosis of gender dysphoria.[31] This was an unusual stance for government officials to take given statistics revealed by the government consultation on reforms to the 2004 Act conducted two years prior revealing that 64 per cent supported removing the requirement for a diagnosis of gender dysphoria.[32] The consultation identified that the number of successful applicants for GRCs was unexpectedly low; since the system's introduction only 4910 trans people had obtained a GRC, out of an estimated UK trans population of 250,000.[33] GEO officials had addressed a submission to the Secretary of State (SoS)[34] suggesting an alternative safeguard to align the medical requirement with current WHO guidelines, by utilising the term 'gender incongruence', which is internationally understood and less stigmatised.[35] However, the SoS replied that such a change would create 'confusion and uncertainty amongst clinicians',[36] and that a medical element to the GRC process ensured appropriate checks and support for applicants. However, the SoS also recognised 'gender dysphoria' as a pathologising term and asserted a keenness to move away from this.[37] The 'final' draft government response to the consultation proposed removing the requirement for a diagnosis of gender dysphoria, replacing it with gender incongruence.[38] However, in September 2020, the SoS back-pedalled, publishing a written Ministerial Statement that the correct balance is struck in the 2004 Act, citing proper checks and balances. Crispin Blunt, a Conservative MP, described this as a 'crushing disappointment' to trans people.[39] Similarly, Marsha de Cordova, Shadow Secretary of State for Women and Equalities, stated that the Government had 'disgracefully let the trans community down'.[40]

30 Ibid [81(a)].
31 Ibid [82].
32 Government Equalities Office, *Reform of the Gender Recognition Act – Government Consultation* (Consultation Paper 2018).
33 *JR111* (n 4 above) [53].
34 Ibid [61].
35 Ibid [62].
36 Ibid [66].
37 Ibid [68].
38 Ibid [71].
39 'Crispin Blunt criticises government trans rights stance' (*BBC News*, 24 September 2020).
40 'GRA: De Cordova and Truss on care for trans people' (*BBC News*, 23 September 2020).

Whittle stated that, when the 2004 Act was drafted, it was an offer that the trans community could not refuse as their options were 'something or nothing'.[41] However, as Hilsenrath highlights, modernising, reviewing and simplifying the process in light of contemporary attitudes is important.[42] Scoffield J assessed materials provided to him regarding Scotland's consideration of a new model of self-declaration for the process of obtaining a GRC. However, as COVID-19 placed the planned reforms on hold, the judge derived little assistance from these, given that the final proposals and outcome remain unknown.[43]

Strasbourg and domestic authority

The court noted that no European consensus exists on the inappropriateness of requiring a psychiatric diagnosis as a condition for gender recognition.[44] Whilst the Parliamentary Assembly adopted Resolution 2048 (2015) on discrimination against transgender people in Europe, calling on member states to, among other things, abolish a mental health diagnosis as a legal requirement to recognise a person's gender identity, member states were permitted wide discretion.[45]

Scoffield J examined pertinent cases in this area, noting the 2004 Act was a response to the European Court of Human Rights (ECtHR) judgment in *Goodwin v United Kingdom*.[46] This case decreed that, as no significant factors of public interest existed to weigh against the interest of the individual applicant in obtaining legal recognition of their gender re-assignment, the fair balance, inherent in the Convention, favoured the applicant, recognising a failure to respect her right to private life in breach of article 8.[47] However, *Goodwin* was a different case because the applicant transitioned after surgery. A more significant hurdle for the present applicant's case emerged from Scoffield J's consideration of the ECtHR decision in *AP, Garçon and Nicot v France*.[48] It was held, *inter alia*, that a requirement to demonstrate the existence of a gender identity disorder in order to secure legal gender recognition did not

41 S Whittle, 'The opposite of sex is politics – the UK Gender Recognition Act and why it is not perfect, just like you and me' 15(3) (2007) Journal of Gender Studies 267, 269.

42 R Hilsenrath, 'Reform of the Gender Recognition Act' (Equality and Human Rights Commission, 16 July 2020).

43 *JR111* (n 4 above) [89].

44 Ibid [97].

45 Ibid [98].

46 [2002] 2 FLR 487.

47 *JR111* (n 4 above) [93].

48 [2017] ECHR 338.

violate article 8.[49] Despite aspects of the government's argument being 'not wholly persuasive', Scoffield J accepted that the requirement for a gender identity diagnosis was aimed at safeguarding the interests of those concerned.[50] Citing the wide margin of appreciation afforded to member states, Scoffield J described this as a powerful submission.[51] He noted Lord Mance's assertion, in *D v Commissioner of Police of the Metropolis,*[52] that there are cases 'where the English courts can and should, as a matter of domestic law, go with confidence beyond existing Strasbourg authority'.[53] Such Convention scrutiny by domestic courts was undertaken in *Carpenter v Secretary of State for Justice*.[54] On the surface, this hindered the applicant as it was held that providing a medical report detailing treatment was not incompatible with the ECHR. However, Scoffield J found from this case that the adequacy of the state's criteria for recognising gender was a justiciable matter. The test then was whether the impugned provisions of the 2004 Act struck a fair balance between the competing interests of the individual and the community as a whole.[55]

Assessment of the fair balance

Requirement for a diagnosis

Scoffield J was satisfied that requiring a relevant diagnosis in support of an application for a GRC remained within the discretionary area of judgment available to Parliament[56] and that, ultimately, this case was not an appropriate platform to '*forge ahead*' of Strasbourg jurisprudence,[57] demonstrating, as Masterman has observed, that British courts are hesitant to develop the meaning of Convention rights.[58] In addressing whether the 2004 Act strikes a fair balance between the needs of the applicant and the community, the judge favoured the respondent's motivations in deterring vexatious applications or abuse of the GRC process, and to provide appropriate support and safeguards for applicants, overall being more consistent with the ECtHR's ruling in *AP, Garçon and Nicot*.[59]

49 *JR111* (n 4 above) [90].
50 Ibid [108].
51 Ibid [110].
52 [2018] UKSC 11.
53 *JR111* (n 4 above) [113].
54 [2015] EWHC 464.
55 *JR111* (n 4 above) [118].
56 Ibid [131].
57 Ibid [132] (emphasis added).
58 R Masterman, T*he Separation of Powers in the Contemporary Constitution* (Cambridge University Press 2011) 203.
59 *JR111* (n 4 above) [133].

Scoffield J accepted the respondent's submission that the legal change in a person's gender is a significant change in their status with potentially far-reaching consequences for them and others, including the state.[60] He agreed that the court's role is not to assess whether the current process is the best or most appropriate way to provide for gender recognition, but noted the 'woefully low' uptake of the GRC process as an indication that the present system is not serving well those it was devised to benefit.[61] However, he maintained that the possible impacts of de-coupling the medical from the legal transition process are matters not well suited to judicial adjudication. Nicol supports this, asserting that courts ought not to compel Parliament if it does not wish to introduce a less stringent process for obtaining a GRC.[62] Scoffield J favoured recognising that there is plainly a medical aspect to some elements of gender transition, at least for some individuals, and, thus, requiring some medical diagnosis is fair.[63]

The required 'disorder' diagnosis

Scoffield J ultimately found the gender dysphoria diagnosis requirement an unnecessary affront to the dignity of a GRC applicant.[64] He cited the English Court of Appeal decision in *R (Elan-Cane) v Home Secretary*,[65] which held that little can be more central to an individual's private life than gender. No reason provided by the respondent adequately explained why recognition should be conditional on proving the existence of a disorder, particularly in light of the development of the international classifications. Scoffield J emphasised that the changes in ICD-11 had not occurred, and thus were not considered by the ECtHR in *AP, Garçon and Nicot*.[66] Further, he found difficulty accepting that specialists could not readily adapt to a similar amendment in the 2004 Act.[67] A 2020 GEO briefing note supported this, stating that the change in terminology 'is largely symbolic and will not interfere with existing clinical processes'.[68] Scoffield J stated that the importance of such symbolism should not, however, be underestimated.[69] The government's analysis of the 2018 consultation responses also noted that, 'a diagnosis of gender dysphoria *or incongruence* is also required

60 Ibid [135].
61 Ibid [137].
62 Cf Nicol (n 22 above) 480.
63 *JR111* (n 4 above) [139].
64 Ibid [140].
65 [2020] 3 WLR 386, [46]–[47].
66 *JR111* (n 4 above) [141].
67 Ibid [144].
68 Ibid.
69 Ibid [145].

in order to access NHS treatment'.[70] This shows gender incongruence is a term known and used by relevant practitioners.[71] A 2018 position statement published by the Royal College of Psychiatrists, specifically recommended 'at the earliest opportunity, de-classify[ing] any terms ... to describe transgender as a mental health disorder'.[72] Pertinently, this professional medical body is responsible for many practitioners likely to provide diagnoses for GRC applications. Scoffield J remarked that the decision to leave the Act untouched appeared to originate from something beyond concern about clinicians coping with a terminology change. Hilsenrath credits this partly to the divisive nature of current debate in this area of law, causing some to withdraw from engaging with discussions.[73] Ultimately, the requirement that diagnosis specifically and expressly be defined as a 'disorder' was ruled not to amount to 'proper checks and balances'.[74] Parliament has been inactive in this area since 2004, and, with today's rapid changes in values militating against an unduly restrictive approach, Scoffield J chose to note that, while the legislature exists to reflect the democratic will of the majority, the judiciary exists to protect minority interests. Parliament is not an expert on the particular diagnostic classifications involved but should be viewed as the arbiter of what safeguards ought to be in place.[75]

Scoffield J concluded that, while the submissions under article 14 added little to the applicant's claims under article 8, the specific requirement of a disorder diagnosis is now unnecessary, unjustified and breached the applicant's article 8 rights. Even with Parliament's discretionary area of judgment and the legitimate aims which the requirement for medical input pursues, the requirement fails to strike a fair balance between the interests of the applicant and those of the community generally.[76] The court held that it would hear further submissions from the parties regarding an appropriate remedy following the decision. This was specifically in relation to the question of whether the legislation could be 'read down' under section 3 of the Human Rights Act 1998, or a declaration of incompatibility would have to be issued under section 4.

---

70 Daniel King, Carrie Paechter and Maranda Ridgway, *Gender Recognition Act: Analysis of Consultation Responses* (Government Equalities Office, CP 284 2020) 41 (emphasis added).

71 *JR111* (n 4 above)[144].

72 Royal College of Psychiatrists, *Supporting Transgender and Gender-Diverse People* (Position Statement, PS02/18, 2018).

73 Cf Hilsenrath (n 42 above) 2.

74 *JR111* (n 4 above) [146].

75 Ibid [147].

76 Ibid [157].

CONCLUSION

In summation, this judgment has significant implications for the process by which transgender people have their acquired gender recognised and marks a step towards better respecting the integrity of transgender people in the UK, tackling one area where they continue to face stigma. Despite what may superficially seem a minor change in terminology from 'gender dysphoria' to 'gender incongruence', as Scoffield J emphasised, the symbolism should not be underestimated, as one word and definition can shift an entire narrative about what it means to be transgender, an aspect of human variation, not a disorder.

Articles

Commentaries and Notes