Winter Vol. 73 No. 4 (2022)

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

Winter Vol. 73 No. 4 (2022)

Contents

Articles

Commentaries and Notes

Northern Ireland
Legal Quarterly
Winter Vol. 73 No. 4 (2022) 587–617
Article DOI: 10.53386/nilq.v73i4.982

Justifying justiciability: healthcare resource allocation, administrative law and the baseline judicial role

Edward Lui*

Magdalen College, University of Oxford
Correspondence email: long.lui@magd.ox.ac.uk

ABSTRACT

A long-standing problem in United Kingdom law concerns the proper relationship between judicial review and healthcare resource allocation. Traditionally, decisions concerning healthcare resource allocation are non-justiciable. This position has already been departed from in the positive law, but few within the academic literature have discussed the theoretical justification for such a departure. This article draws upon the literature on public law theory and makes three theoretical arguments in favour of this departure. First, the doctrine of non-justiciability is an inflexible – and thus inappropriate – form of judicial restraint. Second, one cannot sensibly distinguish cases with an allocative impact (which are justiciable) from decisions concerning healthcare resource allocation. The latter therefore should not be non-justiciable. Third, the *ultra vires* theory entails that decisions concerning healthcare resource allocation should be both justiciable and consistent with the requirements of the rule of law – such that these decisions must be subject to the possibility of both procedural and rationality review. This establishes a baseline judicial role in healthcare resource allocation.

Keywords: administrative law; judicial review; National Health Service (NHS); non-justiciability; rationing; resource allocation.

INTRODUCTION

This article relates to a significant question in United Kingdom (UK) medical and administrative law. The question is this: to what extent, if at all, should decisions concerning healthcare resource allocation in the National Health Service (NHS) be subject to judicial review? In the 1990s, the prevailing judicial view was that the matter is non-justiciable.[1] Healthcare resource allocation is a 'political' matter which should be addressed outside the courtrooms, and preferably in the chambers of Parliament. To this, the court will not – and shall not

* I would like to thank Anne Davies and the anonymous reviewers. Any errors are entirely mine.

1 See eg *R v Cambridge Health Authority, ex p B* [1995] 1 WLR 898, 906.

– intervene.[2] As we shall see, this all is in the past: the doctrine of non-justiciability no longer holds good in the positive law.

This article seeks to provide a theoretical account that underpins the departure from non-justiciability. It will be contended that this change in the positive law is amply justified, by reference to the literature on public law theory. Three arguments will be made to this effect. First, the doctrine of non-justiciability is an inappropriate form of judicial restraint, due to its lack of flexibility. Even if judicial restraint is justifiably called for, the doctrine of non-justiciability is not a suitable means to achieve this. Second, one cannot sensibly distinguish cases with an allocative impact (which are without more justiciable) from decisions concerning healthcare resource allocation. Therefore, the latter – like the former – should be justiciable. Third, the *ultra vires* theory tells us that the power of decision-makers – even for those responsible for healthcare resource allocation – is necessarily limited. Parliament intends that this limitation be maintained by way of judicial review, and so these decisions must be justiciable. One can go even further: since the *ultra vires* theory requires the decision-makers to act compatibly with the rule of law, we can appeal to the requirements of the rule of law to ascertain what principles of administrative law must (at minimum) exist. Through this line of reasoning, we can see why both procedural *and* rationality review must be available to challenge a decision concerning healthcare resource allocation.[3] This discussion establishes a baseline judicial role concerning healthcare resource allocation in UK law.

To those well-versed with this area of medical and public law, it may be intuitively questioned how a piece on the doctrine of non-justiciability would further contribute to the academic literature. A number of experts on this area – such as Newdick, Syrett and Wang – have already noted the departure from non-justiciability in the positive law: the courts have no longer seen the issue of healthcare resource allocation as non-justiciable and have conducted judicial review upon

2 *R v Secretary of State for the Environment, ex p Nottinghamshire County Council* [1986] AC 240, 247.

3 By procedural and rationality review, this article is referring to the two well-known grounds of review in administrative law (as referred to by Lord Diplock in *Council of Civil Services Unions v Minister for the Civil Service* [1985] AC 374) that an applicant may invoke – as opposed to the kinds of remedies that should be available to the applicant, should a challenge be successful. The latter will also be discussed below.

this premise.[4] It may thus seem like a piece examining the issue of non-justiciability is all but tackling an issue of the past.

But this would be painting too simple a picture. Whilst the positive law has undoubtedly moved on, most of the academic literature has focused on what led to the legal development (eg the social/institutional context of the NHS, or the emergence of a 'culture of justification' in UK public law).[5] These contributions are of course valuable, but they leave the question of theoretical justification – ie *whether* (and, if so, *why*) this move is theoretically justified – largely unanswered. Naturally, we academics should not be satisfied with the statement that 'the positive law has moved on', nor should we be satisfied with just knowing that 'factor X has caused the positive law to move on': we must further engage in an inquiry of *whether* this departure is justified. As is well known (and as will be discussed later), the famous case of *B* provides us with one of the most authoritative statements on why courts should maintain a stance of non-justiciability.[6] The clear changes in the positive law since the days of *ex p B* implied disagreement with this authoritative statement by Lord Bingham MR. But few judges and academics have sought to provide an academic exposition of this disagreement. This gap in the literature calls for a piece that engages in public law theory, so as to provide the relevant theoretical grounding for this well-documented change in positive law. This is what this contribution seeks to achieve.

Looking beyond the area of medical law, the litigation in this area has also received attention from public lawyers. It is interesting to note that the public law literature does not seem to have caught up with the developments in this area. In 2007, King described the area of healthcare resource allocation as an area where the doctrine of non-justiciability remains potent.[7] But *even* at the time – as those familiar with this area have noted – the courts had already begun with a more active role in judicial review.[8] Nevertheless, in the latest edition of

4 Christopher Newdick, *Who Should We Treat? Rights, Rationing, and Resources in the NHS* 2nd edn (Oxford University Press 2005) 100–104; Keith Syrett, *Law, Legitimacy and the Rationing of Health Care: A Contextual and Comparative Perspective* (Cambridge University Press 2007) 164–177; Daniel Wei L Wang, 'From *Wednesbury* unreasonableness to accountability for reasonableness' (2017) 76 Cambridge Law Journal 642, 644–649; Keith Syrett, 'Healthcare resource allocation in the English courts: a systems theory perspective' (2019) 70 Northern Ireland Legal Quarterly 111, 114–116.

5 See eg Newdick (n 4 above) 94–109; Wang (n 4 above) 642, 653–656; Syrett, 'Healthcare resource allocation' (n 4 above) 117.

6 *B* (n 1 above) 906.

7 Jeff A King, 'The justiciability of resource allocation' (2007) 70 Modern Law Review 197, 199–200.

8 Newdick (n 4 above) 93–105; Syrett, *Law, Legitimacy and the Rationing of Health Care* (n 4 above) 172–177.

De Smith's Judicial Review – which was published in 2018 – King's piece and *ex p B* have still been cited with approval as representative of the current law. The statement of law remained that 'the allocation of resources is regarded as a matter which is not normally amenable to judicial review'.[9] This represented an alarming – and fundamental – disconnect between the literature on medical law and public law.

This piece, therefore, is not nugatory. It makes two direct contributions to the literature across two important fields of law. First, this piece contributes to the medical law literature by justifying – through public law theory – a development that has been thoroughly noted, but clearly undertheorised. Second, this piece contributes to the public law literature by bridging its disconnect with the medical law literature. This piece will therefore be interesting to not only medical lawyers: but also public lawyers that are also invested in the issue of resource allocation.

Before we proceed any further, there are three caveats that should be noted concerning the intended scope of this article. First, this article is only concerned with the judicial role in healthcare resource allocation insofar as UK law is concerned. Although the comparative literature on health litigations beyond the UK can be helpful to this exploration, this article is not crafted with other jurisdictions in mind;[10] nor is there an implied suggestion that the approach in UK law should be taken in other jurisdictions.[11] Second, the reference to public law theory here refers to a mixture of theoretical contributions from *both* administrative law and constitutional theory in the UK law literature (both of which will be discussed extensively below). The *ultra vires* debate is a prominent debate in the administrative law literature, whereas the literature on the rule of law and judicial restraint are two well-known facets of the constitutional theory literature. The aim here is to unite the precepts arising from these contributions with the judicial treatment of

9 Lord Woolf, Sir Jeffrey Jowell QC, Catherine Donnelly, Ivan Hare QC and Joanna Bell, *De Smith Judicial Review* 8th edn (Sweet & Maxwell 2018) [5-150].

10 That is, this article is simply focused on the UK domestic law – although other jurisdictions can remain relevant to the analysis.

11 A suggestion that has been made elsewhere: see eg Colleen M Flood and Aeyal Gross, 'Litigating the right to health: what can we learn from a comparative law and health care systems approach' (2014) 16 Health and Human Rights Journal 62, 66–67, 69; Daniel Wei L Wang, 'Right to health litigation in Brazil: the problem and the institutional responses' (2015) 15 Human Rights Law Review 617, 640–641. If such a move is to be taken, Syrett and Newdick's contributions (see Keith Syrett, 'Evolving the right to health: rethinking the normative response to problems of judicialization' (2018) 20 Health and Human Rights Journal 121; Christopher Newdick, 'Can judges ration with compassion? A priority-setting rights matrix' (2018) 20 Health and Human Rights Journal 107) may provide a good starting point. This matter will, however, be beyond the scope of this article, as we are only concerned here with UK law alone.

healthcare resource allocation. Third, this article only seeks to justify a *baseline* for the judicial role in healthcare resource allocation: that there must be both procedural and rationality review of such decisions, and that the doctrine of non-justiciability cannot be adopted. This does not (nor is it intended to) preclude further debate on the particular form and intensity of review which courts should apply,[12] including the possibility of judicial deference.[13]

THE DOCTRINE OF NON-JUSTICIABILITY: SETTING THE SCENE

To the readers that are less familiar with this area of law, it will be important to first canvass the material developments in the positive law – before we proceed to justify them. It will be convenient to – as Newdick and Wang did – divide the case law into two batches.[14] The first batch of judicial challenges against resource allocation in the NHS reveals a very high degree of judicial restraint. Take *Hincks*.[15] The applicants applied to the court, complaining that the health services in the area were insufficient. Lord Denning MR rejected the application. He noted the 'grievances which many people feel nowadays about the long waiting list to get into hospital'.[16] But he went on to say '[s]o be it. The Secretary of State says that he is doing the best he can with the financial resource available to him: and I do not think that he can be faulted in the matter.'[17] Or take *Collier*.[18] There the court faced a challenge against the NHS's refusal to conduct life-saving surgery on a child, allegedly because no bed in the intensive care unit was available. The application was swiftly dismissed. Stephen Brown LJ noted that 'even assuming … there is immediate danger to health … [t]his court is in no position to judge the allocation of resource by this particular

12 For it is one thing to ask whether there should be rationality review, and another to ask what intensity with which it should be conducted: see Paul Craig, 'Judicial review, methodology and reform' [2022] Public Law 19, 25–26.

13 For which it may be said that there must be a baseline intensity of review: see Cora Chan, 'Proportionality and invariable baseline intensity of review' (2013) 33 LS 1. But the question with which this article is concerned is even logically prior to this: the argument is for the (baseline) existence of review, and not the (baseline) intensity of it.

14 Newdick (n 4 above); Wang (n 4 above); cf Syrett 'Healthcare resource allocation' (n 4 above) 114.

15 *R v Secretary of State for Health, West Midlands Regional Health Authority and Birmingham Area Health Authority, ex p Hincks* [1980] 1 BMLR 93.

16 Ibid 96.

17 Ibid.

18 *R v Central Birmingham Health Authority, ex p Collier* (unreported, 6 January 1988, Court of Appeal).

health authority ... The courts of this country cannot arrange the lists in the hospital.'[19] The most widely known instance of non-justiciability is perhaps *B*. There a child suffered from acute leukaemia. The health authority refused to fund the proposed treatment of chemotherapy and bone-marrow transplant for the child – which could potentially save her life – partly on the ground that it was not an appropriate use of public funds. Sir Thomas Bingham MR held:

> I have no doubt that in a perfect world any treatment which a patient, or a patient's family, sought would be provided if doctors were willing to give it, no matter how much it cost, particularly when a life was potentially at stake. It would however, in my view, be shutting one's eyes to the real world if the court were to proceed on the basis that we do live in such a world ... Difficult and agonising judgments have to be made as to how a limited budget is best allocated to the maximum advantage of the maximum number of patients. That is not a judgment which the court can make.[20]

There is, however, a caveat here. As Syrett explained, the courts did not *formally* classify decisions on healthcare resource allocation as non-justiciable: 'they retained the capacity to intervene' when the decision was *Wednesbury* unreasonable.[21] But this possibility of intervention never really transpired:[22] as Syrett then added, the courts have applied the rationality review in such a stringent manner that 'allocative decisions were, in effect, *insulated from any judicial scrutiny*, even on procedural grounds'.[23] As Endicott rightly suggested of the first batch of case law, 'the courts will give *practically no protection* against bad decisions in the allocation of a limited budget among competing

19 Ibid.

20 *B* (n 1 above) 906.

21 Syrett (n 4 above) 114. Rationality review is – especially when one looks towards other jurisdictions and the international right to health – not the only way through which judicial scrutiny may take place. For instance, it is possible for courts to use reasonableness and proportionality as such tools: see Katharine G Young, 'Proportionality, reasonableness and economics and social rights' in Vicki C Jackson and Mark Tushnet (eds), *Proportionality: New Frontiers, New Challenges* (Cambridge University Press 2017) 249–250, 252–259. But since we are only concerned with the UK law context here – and challenges to healthcare resource allocation decisions in this context have almost always been made through administrative law, rather than human rights law (see Flood and Gross (n 11 above) 67) – we will focus thereafter on rationality review alone. This is not necessarily the case for other contexts of resource allocation, where human rights law can play a more important role: see eg Ellie Palmer, *Judicial Review, Socio-Economic Rights and the Human Rights Act* (Hart 2007) ch 5.

22 Wang (n 4 above) 653.

23 Syrett (n 4 above) 'Healthcare resource allocation' 114 (emphasis added). See also Palmer (n 21 above) 162, 164–165.

needs'.[24] This led to what in practice was a doctrine of non-justiciability over decisions concerning healthcare resource allocation: *no* applicant, however aggrieved and wronged, will receive protection from judicial intervention. (This position is, however, not true, as we shall see, in respect of the second batch of case law.)

In other words, whilst the court did not apply a *de jure* doctrine of non-justiciability (whereby judicial review is in principle ruled out), it still applied a *de facto* doctrine of non-justiciability (whereby judicial review is in principle available, but is in practice ruled out). When the later analysis referred to 'non-justiciability', it was meant to refer to this *de facto* doctrine of non-justiciability. To this article, however, this distinction is not a material one. Why is that so? For if one believes that judicial review exists to uphold a certain value – eg accountability[25] or the rule of law[26] – this value will be lost if judicial review becomes unavailable. On this count, it will not matter whether its unavailability is *de jure* and *de facto*: the value secured by judicial review will still be lost if courts hold that judicial review remains in principle available, but that it be only available on grounds that can never be established in practice. Similarly, if there are arguments against the unavailability of judicial review (eg that certain undesirable consequences follow from the unavailability of judicial review), the validity of these arguments is naturally predicated on the premise that judicial review is unavailable. But again, this premise can be established insofar as judicial review is indeed unavailable – whether this is proven by way of a doctrine of *de jure* or *de facto* non-justiciability. It makes no difference *how* judicial review is rendered unavailable: it only matters here that it *is* indeed rendered unavailable.

This doctrine of non-justiciability can be readily seen from the academic commentary on the abovementioned cases. James and Longley contended that the judgment in *B* 'did little to move substantive review in this area on from the earlier cases of *ex parte Collier* ... which had been notable only for their lack of perceptive analysis and the ritual invocation of *Wednesbury* principles'. The decisions 'in essence not only gave health authorities a free hand to allocate resources as they chose, but also weakened the potential role of the courts'.[27] As mentioned earlier, Wang separated the authorities on reviewing NHS

24 Timothy Endicott, *Administrative Law* 5th edn (Oxford University Press 2021) 272 (emphasis added).

25 Paul Craig, 'Accountability and judicial review in the UK and EU: central precepts' in Nicholas Bamforth and Peter Leyland (eds), *Accountability in Contemporary Constitution* (Oxford University Press 2013) 185.

26 See eg T R S Allan, *Law, Liberty, and Justice: The Legal Foundations of British Constitutionalism* (Clarendon Press 1994) ch 8.

27 Rhoda James and Diane Longley, 'Judicial review and tragic choices: *Ex parte B*' [1995] Public Law 367, 371–372.

resource allocation decisions into two stages. The first stage concerns the authorities that have already been referred to – such as *Hincks* and *Collier*. In those cases,

> the courts restrained themselves to a minimal level of scrutiny of the allocative choices and trusted primary decision-makers to make the best decisions. The court's reasoning was straightforward: resources are scarce, not all health needs can be met, and thus rationing is necessary; and health authorities are best able to do this.[28]

The doctrine of non-justiciability did not subsist. The second batch of cases on resource allocation decisions in the NHS reveals that the courts have moved onwards. It has already been mentioned that this article is not the first to make this observation. So the aim here is only to briefly canvass this move in preparation for the theoretical account in the next section. There are a number of cases within this batch: but it will be quite unnecessary to go through each of them here. The aim here is only to illustrate the legal development with two cases: *A, D and G*[29] and *Otley*.[30]

In *A, D and G*, the applicants were patients suffering from gender identity dysphoria. They applied for funding for their treatment – which included gender reassignment surgery. The health authority refused. It did so on the basis of its policy. It said that if a treatment was regarded as 'clinically ineffective' – and gender reassignment surgery was regarded as such a treatment – it would be accorded a 'low priority'. The aim of this policy was to ensure that the resources of the NHS were 'used appropriately'.[31] No funding would be provided, unless there was an 'overriding clinical need or exceptional circumstances'.[32] The health authority did not find that the applicants had satisfied such criteria. In return, the applicants' contention was that this policy was irrational.

Auld LJ proceeded differently. He first applied *B* and held that '[t]he precise allocation and weighting of priorities is clearly a matter of judgment for each authority'.[33] This was unobjectionable, as the court should not substitute the judgment (as opposed to conducting a rationality review) of the resource allocation decision. But instead of applying the doctrine of non-justiciability, he proceeded to find the policy irrational. This finding of irrationality was not done on the basis of challenging the *medical* assessment by the decision-maker –

---

28 Wang (n 4 above) 643, 645–646.

29 *R v North West Lancashire Health Authority, ex p A, D and G* [1999] EWCA Civ 2022, [2000] 1 WLR 977.

30 *R (Otley) v Barking and Dagenham NHS Primary Care Trust* [2007] EWHC 1927 (Admin).

31 *A, D and G* (n 29 above) 983.

32 Ibid 984.

33 Ibid 991.

for which the health authority clearly possesses relative institutional competence. Thus, Auld LJ held that '[i]n my view, a policy to place transsexualism low in an order of priorities of illnesses for treatment and to deny it treatment save in exceptional circumstances such as overriding clinical need is not in principle irrational'.[34] The flaw of the policy was that it was illogical. The health authority recognised that gender identity dysphoria *was* an illness. But it did not recognise gender reassignment as an effective treatment: and so *no* overriding clinical need could ever have been recognised. The problem therefore was not a flawed medical judgment, but one of logic: if the policy purported to identify exceptions, it must do so 'genuinely'.[35] In this case, the policy was applied in a way which *de facto* imposed a blanket ban on funding for gender identity dysphoria: and the exception was therefore a pretence. This shows that the court is willing to carefully scrutinise the nature of the particular issue thrown up in the case: a decision on healthcare resource allocation may be illogical on grounds that can be readily scrutinised by the court, although it may not possess medical expertise which matches that of the health authorities.

The second case is *Otley*. The applicant suffered from metastatic colorectal cancer. She had received chemotherapy previously, but her body's response was poor – so poor, indeed, her doctor found the treatment 'of absolutely no value'.[36] She discovered the drug Avastin in her own capacity, and self-funded five rounds of it. Her body's response was excellent. She then applied to the NHS for funding further Avastin treatment. According to NHS policy, treatment with Avastin would not be funded unless there were exceptional circumstances. This was because Avastin was not regarded as a sufficiently cost-effective drug. It was found that the applicant did not fit the exceptionality criteria, and so funding was refused. The challenge is founded on the basis that it was irrational for the health authority not to regard the applicant as an exceptional case.

Just as in *A, D and G*, the doctrine of non-justiciability did not apply. Mitting J held in favour of the applicant. He found that 'on any fair minded view of the exceptionality criteria … [the applicant's] case was exceptional'.[37] The query, again, is not one based on any competing medical expertise asserted by the court. The matter, rather, is one in which the court is sufficiently competent to make a fair assessment. One can see that from the applicant's contention:

> [The applicant] was at the time when the decision was made, as she had been throughout, relatively fit. She was young by comparison with

34 Ibid.
35 Ibid.
36 *Otley* (n 30 above) [2].
37 Ibid [26].

> the cohort of patients suffering from this condition. Her reactions to other treatment … had been adverse. Her specific clinical history suggested that … Avastin had been of benefit to her. By comparison with other patients, she, unlike many of those the subject of the studies, had suffered no significant side-effects from a cocktail which included Avastin.[38]

The question, then, was whether the applicant's circumstances were such that her case was exceptional. The court is not required to substitute the judgment of the decision-maker. It need only see whether the judgment is irrational. It will do so, logically, by comparing the usual case of a patient requesting Avastin treatment to the applicant's case. It is true that the judge cannot perform a medical assessment of the applicant, or readily understand the scientific studies on the drug to comprehend its usual effectiveness. But, as Chan noted, it must be remembered that, although the court may be generally less competent than a decision-maker, the deficit of institutional competence can potentially be remedied. An information gap, for instance, can be addressed by a disclosure of information or expert evidence.[39] In *Otley*, this was the basis upon which the court found itself in an appropriate position to interfere: the court may proceed on the basis of the applicant's expert evidence to assess if the NHS decision was properly made. The theme here is similar to that in *A, D and G*: although we are faced with a decision concerning healthcare resource allocation, it may remain true that the court possesses sufficient competence to assess the rationality of the decision. If so, the doctrine of justiciability would appear to be too excessive a response to the needs of judicial restraint.

Pausing here, *A, D and G* and *Otley* are highly instructive on two counts. First, they have demonstrated to us – as a matter of legal reality – that the courts have no longer applied the doctrine of non-justiciability. This is despite the fact that cases such as *Collier* and *B* have never been formally overruled. Second, they have shown us that the issues thrown up in a judicial review of decisions concerning healthcare resource allocation can be ones that courts are relatively competent to assess (this will be further developed in the next section).

38 Ibid [20].

39 Cora Chan, 'A principled approach to judicial deference for Hong Kong' in Guobin Zhu (ed), *Deference to the Administration in Judicial Review: Comparative Perspectives* (Springer 2019) 217.

THE THEORETICAL CASE FOR JUSTICIABILITY

As has been mentioned in the introduction to this piece, there is a critical gap in the literature on this important legal development: few have attempted to inquire whether (and if so, why) this development is theoretically justified. This piece aims to fill this gap. As Syrett noted, the standard case *for* non-justiciability – adopted by the judges in the first batch of cases – is two-pronged.[40]

a. *Institutional competence*: courts are not sufficiently competent to adjudicate on issues of healthcare resource allocation. Various (sometimes overlapping) reasons for this conclusion have been advanced. Judges are only legally trained (with little background in say healthcare economics and healthcare management),[41] whereas healthcare resource allocation can engage complex scientific, political and moral issues.[42] The adversarial nature of the proceedings may make it difficult to gather information comprehensively, especially when the litigants before the court may not represent the general run of patients[43] and given the polycentricity involved.[44] Courts are also compelled to apply

40 Syrett, *Law, Legitimacy and the Rationing of Health Care* (n 4 above) ch 5. This point can also be seen in many of the pieces cited within this section.

41 Ibid 131, 144–145; Daniel Wang and Benedict Rumbold, 'Priority setting, judicial review, and procedural justice' in Andelka M Phillips, Thana C de Campos and Jonathan Herring (eds), *Philosophical Foundations of Medical Law* (Oxford University Press 2019) 187–189; Woolf et al (n 9 above) [5-150]; Stefanie Ettelt, 'Access to treatment and the constitutional right to health in Germany: a triumph of hope over evidence?' (2020) 15 Health Economics, Policy and Law 30, 32. See further Leticia Morales, 'Judicial interventions in health policy: epistemic competence and the courts' (2021) 35 Bioethics 760, 761–763, but also 764–765.

42 Wang (n 11 above) 635–636.

43 Christopher P Manfredi and Antonia Maioni, 'Courts and health policy: judicial policy making and publicly funded health care in Canada' (2002) 27 Journal of Health Politics, Policy and Law 213, 218–219; Ettelt (n 41 above) 32, 36–38; Syrett, *Law, Legitimacy and the Rationing of Health Care* (n 4 above) 130, 147–148. There can be further problems. The litigants may seek to adduce evidence that favours their predetermined conclusion – whilst ignoring evidence pointing to the contrary direction: see Susan Haack, 'What's wrong with litigation-driven science? An essay in legal epistemology' (2008) 38 Seton Hall Law Review 1072, 1077. The litigants may also conduct their cases in a way that focuses only on one aspect of the health policy (particularly if this may favour this case), without presenting the whole picture to the court: see Manfredi and Maioni (ibid) 222, 228. This can make it even more difficult for the court to impartially and comprehensively assess all the relevant evidence that concerns health policy.

44 Wang (n 11 above) 630, 636; Ettelt (n 41 above) 32; Daniel Wei L Wang, 'Priority-setting and the right to health: synergies and tensions on the path to universal health coverage' (2020) 20 Human Rights Law Review 704, 723–724.

legal reasoning, which may not be the most suitable for evaluating policy alternatives in an open-minded manner.[45]

b. *Constitutional legitimacy*: in a constitutional democracy, the issue of healthcare resource allocation falls within the province of the political branches of government (ie the executive and legislature). In light of this, courts should not attempt to substitute their own policy preferences on this issue: but they should leave the decision to those that enjoy more democratic legitimacy.[46]

For these two reasons – so the standard case goes – the matter of healthcare resource allocation is not appropriate for determination by courts; that is, it should be non-justiciable.[47] Indeed, the UK courts are not entirely 'out of the woods' yet with the standard case. Although the courts do not necessarily adopt the full form of non-justiciability (as seen in the first batch of cases), similar concerns have still rippled in the second batch of cases – calling on occasions for an acute curtailment of the judicial role beyond procedural matters in healthcare resource allocation, based particularly on concerns for the court's relative lack of institutional competence.[48] So, although the standard case is not now precisely followed, it has continued to play an influential role in the positive law. An examination of its validity is thus particularly apt and important.

The theoretical account here consists of two related, but distinct, propositions: (a) the standard case for non-justiciability (based on institutional competence and constitutional legitimacy concerns) is flawed; and (b), from the perspective of public law theory, all healthcare resource allocation decisions should be justiciable. The departure from the doctrine of non-justiciability is perforce theoretically justified: and (it follows) that there should be a baseline judicial role concerning healthcare resource allocation. Before we proceed any further, one must first be clear about the relationship between these two propositions.

---

45 Manfredi and Maioni (n 43 above) 218, 222, 226, 234. This may be echoing some more generalist concerns, as reflected in Jeremy Waldron, 'Judges as moral reasoners' (2009) 7 International Journal of Constitutional Law 2.

46 Syrett, *Law, Legitimacy and the Rationing of Health Care* (n 4 above) 132; Wang and Rumbold (n 41 above) 187–189; Woolf et al (n 9 above) [5-150]; Ettelt (n 41 above) 32; Wang (n 44 above) 714. See also Pavlos Eleftheriadis, 'A right to health care' (2012) 40 Journal of Law, Medicine and Ethics 268, 282.

47 Syrett, *Law, Legitimacy and the Rationing of Health Care* (n 4 above) 128–129; Woolf et al (n 9 above) [1-040].

48 Keith Syrett, 'Health technology appraisal and the courts: accountability for reasonableness and the judicial model of procedural justice' (2011) 6 Health Economics, Policy and Law 469, 471, 473, 477–480; Wang and Rumbold (n 41 above) 186–187, 190. Nevertheless, one can still see a material distinction between the second batch of cases from the first: see Newdick (n 4 above) 93, 98–102, 105–107; Wang (n 4 above) 643–651.

These two propositions are distinct: the former is a negative case *against* non-justiciability, whilst the latter is a positive case *for* justiciability. It is logically possible – for instance – for a critic to disagree with the standard case for non-justiciability, but to present an alternative case for non-justiciability. In light of this possibility, the negative case alone may not suffice. But even if any alternative case is now to be made, that critic will have to respond to the positive case for justiciability propounded in this piece. This is why the theoretical account consists of two distinct – but mutually reinforcing – propositions.

Against the standard case: inflexibility

The doctrine of non-justiciability is an inappropriate form of judicial restraint, due to its overt inflexibility. The standard case for non-justiciability posits that courts lack institutional competence[49] and constitutional legitimacy[50] compared to the decision-maker in healthcare resource allocation decisions, and the courts must be sensitive to these differences. Let us assume here that the court is generally less capable and constitutionally legitimate than a decision-maker in the context of healthcare resource allocation, as has been suggested by many in the academic literature.[51] The argument here does not deny these concerns, but only doubts whether non-justiciability properly follows from this premise. To say that this *invariably* leads to a doctrine of non-justiciability in healthcare resource allocation (as the first batch of case law suggests) ignores two facts: (a) the extent to which the court lacks institutional competence and constitutional legitimacy can vary; and (b) other factors may also influence the proper extent of judicial restraint. In other words, the doctrine of non-justiciability applies a 'spatial' approach to judicial restraint.[52] This concept arose from the literature relating to judicial deference in human rights cases, but it equally applies here. It means that the courts will carve out 'wholesale subject areas as automatically warranting a small or large degree of deference'.[53] In this case, the degree of judicial restraint is the furthest one can go: non-justiciability. The idea is that whenever

49 E Palmer, 'Resource allocation, welfare rights—mapping the boundaries of judicial control in public administrative law' (2000) 20 Oxford Journal of Legal Studies 63, 76.

50 Martin Chamberlain, 'Democracy and deference in resource allocation cases: a riposte to Lord Hoffmann' [2003] Judicial Review 12 [13]; Keith Syrett, 'Impotence or importance? Judicial review in an era of explicit NHS rationing' (2004) 67 Modern Law Review 289, 295; *R (Pfizer Ltd) v Secretary of State for Health* [2000] EWCA Civ 1566, [2003] 1 CMLR 19 [17].

51 See eg Manfredi and Maioni (n 43 above); Woolf et al (n 9 above) [5-150]; Wang (n 44 above) 723–724.

52 King (n 7 above) 421.

53 Chan (n 39 above) 217.

we have a decision concerning healthcare resource allocation (as the input) the court will automatically proffer non-justiciability (as the output) regardless of other contextual factors. This discussion leads us to two criticisms against the doctrine of non-justiciability, as posited by the standard case and the first batch of case law.

The first criticism is that, even within the area of healthcare resource allocation, the court is not inevitably inapt. While the court may not have a varying democratic mandate, it does have a varying degree of institutional competence compared to the decision-maker. Let us contrast two decisions that we have canvassed earlier: *B* and *A, D and G*. In both cases, the health authorities refused to fund the applicant's treatment by reason of budgetary concerns. But there is nevertheless a critical difference. In *B*, the attack was taken against the balancing of the applicant's individual needs and the authority's financial constraints. The court found itself out of its depth – compared to the health authority – and so applied the doctrine of non-justiciability. In *A, D and G*, the attack was taken against the funding *policy*. The challenge was that the policy was illogical. The policy stated that the funding would only be given in exceptional circumstances. But, at the same time, the doctors in charge did not believe that there could be an effective treatment for gender identity dysphoria (from which the applicant suffered), so the applicant could never have fulfilled the criteria under the policy. The policy is therefore irrational: it purports to provide a policy of exceptionality, whilst in fact it is a 'blanket policy'.[54]

It is not here suggested that *B* is right to apply a doctrine of non-justiciability. But it is suggested that, by contrasting these two cases, one can see how – even within the area of healthcare resource allocation – the court's relative institutional competence is not uniform. In *B*, the challenge was more about the delicate and difficult task of managing resources and balancing various needs to be met by the NHS. In *A, D and G*, the challenge was more about the logicality of the policy. There is no reason why a doctor is in any better a position than a judge to assess this matter. Indeed, a judge – who is experienced in dealing with logic and reasons – would likely be a better expert than a doctor. As Chan suggested, after identifying the issue to be dealt with 'the court should then ask whether it indeed suffers from incompetence thereon. If it is a question of logic or one concerning which the court has all of the relevant information it needs to decide, then the court suffers no institutional incompetence.'[55]

54 *A, D and G* (n 29 above) 993–994; Keith Syrett, 'Rationing, resource allocation, and appropriate medical treatment' in Sara Fovargue and Alexandra Mullock (eds), *The Legitimacy of Medical Treatment: What Role for the Medical Exception?* (Routledge 2015) 207–208.

55 Chan (n 39 above) 220.

The second (related) criticism is this: by taking a formalist approach, the doctrine of non-justiciability ignores matters which may be relevant to the appropriate degree of judicial restraint – beyond the fact that the issue of healthcare resource allocation is touched upon. In a separate article, King discussed the various approaches to judicial restraint. There is the non-doctrinal approach – where the judges will 'use their good sense of restraint on a case-by-case basis rather than employ any conceptual framework'.[56] There is the formalist approach – where the courts will 'apply abstract categories' that 'they believe properly allocate decision-making functions between different branches of government'. Prominent ones include 'law', 'politics', 'principle' and 'policy'.[57] There is, finally, the institutional approach to judicial constraint. It 'focus[es] on the comparative merits and drawbacks of the judicial process as an institutional mechanism for solving problems': and in applying it, the courts will weigh the relative institutional competence of the court as a factor towards the degree of judicial scrutiny (eg a balancing stage).[58] The doctrine of non-justiciability is a formalist approach. It labels certain decisions (ie those decisions concerning healthcare resource allocation) as non-justiciable.[59]

This means that the doctrine of non-justiciability is subject to King's arguments against the formalist approach to judicial restraint. There are several of them. First, it is highly rigid and ignores any adverse consequences it produces. This encourages a view that, as long as a decision follows this 'doctrine' – and is hence conceptually correct – 'its consequences are of minor importance'. This may be true in the context of healthcare resource allocation decisions. It may be easy for judges to simply invoke the doctrine of non-justiciability, while ignoring the potentially disastrous consequences (eg death, when one thinks about the NHS context) that this may have on the rejected applicants.[60] Second, it 'obviate[s] the need for ... analysis'. It stops the court from thinking why, in this particular case and with this particular context, it should not intervene.[61] Rather, it will say to itself: since we are concerned with healthcare resource allocation – and so the doctrine of non-justiciability is engaged – this is the end of the matter.[62] But this may not be true. Some cases may feature additional factors – which could well influence the proper approach for the courts to take – such

56 Jeff A King, 'Institutional approaches to judicial restraint' (2008) 38 Oxford Journal of Legal Studies 409, 410–411.
57 Ibid 414–416.
58 Ibid 410, 427.
59 Ibid 420–422.
60 Ibid 414.
61 Ibid 415.
62 Ibid 421.

as if the matter involves highly important interests (eg the survival of the applicant) or rights (eg the right against non-discrimination). The third point is related to, but distinct from, the second: the doctrine of non-justiciability carries with it an uncompromising approach. It deems all decisions concerning healthcare resource allocation non-justiciable, while not recognising that the appropriate degree of judicial restraint may vary.[63] Even assuming that *some* healthcare resource allocation decisions may be properly non-justiciable, it does not mean *all* of them, regardless of context, are. A court that is truly sensitive to the constitutional needs for judicial restraint would recognise that, at times, a lesser degree of judicial restraint would be properly called for. The doctrine of non-justiciability does not fulfil this need.

These two heads of criticism meet up to form the proposition that the doctrine of non-justiciability lacks flexibility. In the literature about deference in human rights litigation, Allan contended that the doctrine of deference is illegitimate, partly because it can 'collapse into a non-justiciability doctrine'.[64] To this Kavanagh retorted:

> Both deference and non-justiciability are based on concerns about the institutional limits of the judicial role when compared to the competence, expertise and democratic legitimacy of the elected branches. This is what makes them similar doctrines. However, deference and non-justiciability also differ in significant ways ... Deference ... is a more flexible doctrine which is not antithetical to judicial scrutiny. There are degrees of deference and establishing the appropriate degree is a matter of balancing all the relevant factors in the individual case. Rather than being a blanket rule preventing scrutiny, deference maintains some flexibility by requiring the courts to assess their institutional competence to deal with a particular issue, and to show restraint to the extent that their competence is limited ... The relative flexibility of the doctrine of deference and the fact that it does not remove certain issues from judicial scrutiny altogether, are the main advantages of deference over non-justiciability.[65]

These words ring equally true here. They demonstrate to us the ineptitude of the doctrine of non-justiciability as a proper approach to exercising judicial restraint, even in the context of healthcare resource allocation. This theoretical discussion can be bolstered by reference to other examples we see in comparative law – where the extent of judicial restraint applied towards scrutinising healthcare resource allocation is more flexibly adjusted. One more well-known example is the South African jurisprudence on the right to health, where the court applies a

63 Ibid 418, 421–422.

64 T R S Allan, 'Human rights and judicial review: a critique of "due deference"' (2006) 65 Cambridge Law Journal 671, 688–689.

65 Aileen Kavanagh, 'Defending deference in public law and constitutional theory' (2010) 126 Law Quarterly Review 222, 244–245.

standard of reasonableness to the impugned decision.[66] This standard does not preclude the need for judicial deference, based on grounds reflected in the standard case for non-justiciability. In the landmark decision of *Soobramoney*,[67] the court held that the court should accord deference to healthcare resource allocation decisions – in light of the difficulty this involves – and expressly cited *B* with approval.[68] But, as Young noted, the reasonableness standard is 'context-driven' and can be much more exacting than the *Wednesbury*[69] standard of review – applying substantive control on government decision-making particularly when the decision affects the more vulnerable sectors of society.[70] This displays flexibility in the exercise of judicial restraint, although the starting point prescribed by *Soobramoney* is based on concerns very similar to the standard case for non-justiciability considered herein (indeed, *B* was itself cited with approval).

Another interesting example is the German right to health. The German right to health guarantees a right to substantive treatment,[71] but it is not an unqualified right that neglects entirely the relevance of cost-effectiveness.[72] The right is 'criteria-based': that only in a limited category of life-threatening cases, the court will require treatment to be provided; and although there still needs to be some clinical evidence in favour of the treatment sought, the threshold to be met is clearly relaxed.[73] This move is a clear response to the need to adjust the proper judicial role, based on the impact of the decision on the individuals affected.[74] Whilst these examples may not necessarily represent the most suitable approaches for UK law, they at least illustrate how

66 *Minister of Health v Treatment Action Campaign* [2002] ZACC 15; 2002 (5) SA 721 [30]–[39], [52], [58]–[59].

67 *Soobramoney v Minister of Health (Kwazulu-Natal)* [1997] ZACC 17, 1998 (1) SA 765.

68 Ibid [19], [29]–[30]; see also *Treatment Action Campaign* (n 66 above) [38].

69 *Associated Provincial Picture Houses Ltd v Wednesbury Corporation* [1947] 1 KB 223.

70 Young (n 21 above) 252–255, 261, 268.

71 Ettelt (n 41 above) 38.

72 Cf Wang (n 11 above) 621–624, 626, 629.

73 Ettelt (n 41 above) 34, 36–40; see also Newdick (n 11 above) 117.

74 Palmer (n 21 above) 188; Mark Elliott, 'From bifurcation to calibration: twin-track deference and the culture of justification' in Hanna Wilberg and Mark Elliott (eds), *The Scope and Intensity of Substantive Review: Traversing Taggart's Rainbow* (Hart 2015) 70, 76–79; Ernest Lim and Cora Chan, 'Problems with Wednesbury unreasonableness in contract law: lessons from public law' (2019) 135 Law Quarterly Review 88, 98–100.

judicial restraint in the context of healthcare resource allocation can be flexibly exercised[75] – short of a doctrine of non-justiciability.

Against the standard case: allocative impact

It cannot be controversial that cases with an 'allocative impact' – ie where 'the effect of the decision is to impose a *financial burden* upon public resources'[76] – *are*, without more, justiciable. This requires some elaboration. When we speak of cases with an allocative impact, the relevant financial burden may come in many forms. They include damages awarded, legal costs, 'costs of administrative compliance' and 'diversion of resources' to avoid future liability and future claims allowed by the judgment.[77] What is common amongst these scenarios is that the court – through rendering a decision – compels the re-allocation of public resources by the authorities.[78] It may be that, after the decision, the costs of a certain government department would gravely increase – as a result of which the central Government may have to reallocate its limited budget, so more funds will go to that department. Or the central Government may decide otherwise: and ask the government department to live with its current budget. Then the reallocation will have to be done within the government department: it may have to cut certain parts of its existing services and staffing, so as to support its increased expenditure. None of this would have occurred but for the judgment. There is, strictly speaking, no court order (as in a mandatory injunction) compelling this resource reallocation. But the reallocation remains, in reality, compelled by the judgment.

Most importantly – subject to a qualification below – decisions with an allocative impact are *inevitable*. Take, by way of example, cases dealing with the liability of the police in negligence to the victims of criminals such as *Hill*[79] and *Michael*.[80] Chamberlain explained

---

75 Indeed, in judicial review relating to other contexts such as social care and taxation, the courts have sometimes acted with a more flexible form of judicial restraint – although similar concerns of institutional competence and constitutional legitimacy may arise: see eg Palmer (n 21 above) 222–224, 233–239; Jeff A King, 'The pervasiveness of polycentricity' [2008] Public Law 101.

76 King (n 7 above) 208 (emphasis added).

77 Ibid 209.

78 Ibid 218. This does not thereby suggest that *whenever* the court makes any decision, it will thereby compel a re-allocation of public resources. (This is due to a qualification that will be addressed below.) Rather, the suggestion is that there are these very commonplace scenarios where a court will compel a re-allocation of public resources – and no one will seriously suggest that these cases should be without more non-justiciable.

79 *Hill v Chief Constable of West Yorkshire* [1989] AC 53.

80 *Michael v Chief Constable of South Wales Police* [2015] UKSC 2, [2015] AC 1732.

that 'whichever way [the court] decide[s]', the decision will have an allocative impact. 'If the police are liable, some resources will have to flow from' other parts of the budget – the intended beneficiary of which could be the other parts of the public – 'to the victims'. If the police are not found liable, 'the result will be that the victims will be denied the resources they would otherwise have had'.[81] One can go even further than Chamberlain's analysis. The decision – whichever way it is decided – does not only affect the victim in the immediate case. It also affects the future victims whose claims will be affected by the ruling (say how the decision in *Hill* will affect the victim in *Michael*). The decision may also influence how the police will in the future conduct themselves. Government policies may be adjusted: if liability will be more stringently imposed, defensive behaviour may occur and more resources may be dedicated towards avoiding future liability. The contrary is also true. On this analysis, it does not really matter how a judicial challenge to healthcare resource allocation is framed – whether by way of legality, procedural, rationality or other grounds of review – the decisions will still inevitably have an allocative impact (which can be unknown, unforeseen and even possibly, unknowable and unforseeable). The same analysis may conceivably be made of many decisions in contract, property and commercial law – but it is not necessary to repeat the analysis once more.[82]

Since cases with an allocative impact are perforce inevitable, it is uncontroversial that this cannot *per se* provide a defence to judicial scrutiny: ie these cases are *not per se* non-justiciable. For otherwise, the judicial role may have to be destroyed altogether. If we proceed from this starting point, we can see how one may come into conflict with the standard case for non-justiciability: for a legal challenge to the decision on healthcare resource allocation is – of course – a case with an allocative impact. So the standard case for non-justiciability must therefore *distinguish* the decisions on healthcare resource allocation (which, the critic says, are non-justiciable) from other cases with an allocative impact (which are justiciable).

One possibility is to invoke King's analysis: he distinguishes cases with an allocative impact from a separate category of cases – which he calls 'discretionary allocative decision-making'. This category of cases is defined as where a decision-maker makes a 'discretionary' decision to allocate public resources, *and* the decision 'take[s] account of the cost of the allocation'.[83] And cases of 'discretionary allocative decision-making' – unlike mere cases with an allocative impact – can be non-justiciable. King further suggests that decisions concerning healthcare

81 Chamberlain (n 50 above) [14].
82 King (n 75 above) 109.
83 King (n 7 above) 197–200.

resource allocation fall within this category, since he expressly cites *B* as an example of 'discretionary allocative decision-making'. In positing the distinction, King is aware that the distinction may come under attack. An argument may be run to the effect that:

> [A] judicial decision causing allocative impact amounts to the same thing as judicial review of discretionary allocative decisions. In both cases the court forces the government to reallocate from one area to another and on an issue that is better decided by the government. If this is the case, then why make the distinction in the first place?[84]

The idea is simply this: if we accept that cases with an allocative impact to be properly adjudicated upon by courts, it seems rather odd to find challenges to 'discretionary allocative decision-making' to be entirely unsuited for adjudication. For, subject to a qualification below, both categories of cases are – after all – about the courts forcing government to reallocate resources. One possible defence for the distinction (says King) is 'that even where precisely the same financial sums are at stake, there may be institutional reasons for allowing the courts to decide legal questions having allocative impact' – but not more than that. While '[i]t may be better for a court to have the decisive say on' legal issues such as 'a statutory duty', 'legitimate expectation', it is 'quite another thing for a court to decide whether a mere one thousand pounds is better allocated to either of two people with putatively similar legal rights'.[85]

If this defence succeeds, it seems like the critic may potentially maintain the standard case for non-justiciability over healthcare resource allocation – whilst accepting cases with an allocative impact to be properly justiciable. But there are at least two possible responses to a defence run on this line. First, this defence posits a formalist distinction between 'legal' and 'non-legal' questions. Yet a critic that uses this reason to justify a bar against rationality review seems to forget that rationality review *is* a legal question. *Wednesbury* is undoubtedly a legal test – begging a question of law – as much as its counterpart[86] proportionality is a question of law.[87] The critic may then seek to draw a distinction between different kinds of questions of law. The critic may say that rationality review is, although a question of law, heavily influenced by political matters like resource allocation.

---

84 Ibid 218.

85 King (n 7 above) 219. See a similar point concerning social rights adjudication in Daniel Wei L Wang, 'Social rights adjudication and the nirvana fallacy' [2018] Public Law 482, 484.

86 These two doctrines being analogous: see Rebecca Williams, 'Structuring substantive review' [2017] Public Law 99.

87 *Kwok Cheuk Kin v Secretary for Constitutional and Mainland Affairs* [2017] HKCFA 44, (2017) 20 HKCFAR 353 [29].

This is different from some more 'purist' questions of law, like statutory interpretation that has an allocative impact. This distinction does not withstand scrutiny. For, first, it seeks to cover a problematic formalist distinction ('legal' and 'non-legal') with another layer of problematic formalist distinction (along the lines of 'legal but political' and 'legal but apolitical'). Second, and most importantly, the critic is seeking to justify not a mere doctrine of deference, but a doctrine of non-justiciability. It is clear, since *Miller (No 2)*,[88] that the fact that a legal question is embroiled in a political context does not mean it is non-justiciable. As Baroness Hale and Lord Reed explained:

> [A]lthough the courts cannot decide political questions, the fact that a legal dispute … arises from a matter of political controversy, has never been sufficient reason for the courts to refuse to consider it … [A]lmost all important decisions made by the executive have a political hue to them. Nevertheless, the courts have exercised a supervisory jurisdiction over the decisions of the executive for centuries.[89]

The second response is this: this defence misunderstands the nature of rationality review. There are two distinct concepts that have been mixed up in this defence, and indeed in the older cases such as *Collier*: (a) the court substituting the resource allocation decision to be made by the decision-maker and (b) the court interfering with resource allocation decisions by applying rationality review. These concepts are familiar ones in the literature concerning *Wednesbury* and proportionality review. Concept (a) refers to the proposition that

> the reviewing court will decide the case de novo as if it had been the primary decision-maker … on this view the court considers the facts, makes its own decision as to what the proportionate outcome should be and does so without giving any particular weight to the primary decision-maker.[90]

Per the defence, the courts are institutionally incompetent to substitute the judgment of the decision-maker. Rationality review in concept (b), however, does *not* involve this. It does not require the courts to directly compete with the expertise of the decision-maker, for the court is not purporting to 'reassess the matter afresh and decide … that funds ought to be allocated in one way rather than another'.[91] On Daly's analysis, the court is simply examining if the 'indicia of unreasonableness' – such as 'illogicality' and 'disproportionality' – exist.[92] It is one thing

88 *R (Miller) v The Prime Minister* [2019] UKSC 41, [2020] AC 373.

89 Ibid [31]. See also John Laws, 'Law and democracy' [1995] Public Law 72.

90 Paul Craig, 'The nature of reasonableness review' (2013) 66 Current Legal Problems 131, 141.

91 Paul Craig, *Administrative Law* 9th edn (Sweet & Maxwell 2021) [21-002].

92 Paul Daly, '*Wednesbury*'s reason and structure' [2011] Public Law 238, 242–247.

to say that the courts do not have the expertise to allocate healthcare resources from scratch – which no doubt is a difficult task particularly for lawyers and can engage concerns of institutional competence[93] – but quite another to say that the courts do no have the expertise to even scrutinise the coherence of the premises and reasoning underpinning the decision altogether. *A, D and G* is an example where the court can quite competently undertake the latter task, without claiming to be able to undertake the former task. For it is fairly possible for parents to criticise a teacher at a primary school, without having all the expertise for pedagogy themselves; the relationship envisaged here between courts and decision-makers in rationality review is similar. Rationality review does not (and cannot) entail the substitution of judgment,[94] just as the parent does not by criticising the teacher thereby take over the teacher's role: and it follows that one may not object to rationality review on the basis that the courts are thereby substituting the judgment of the decision-maker.

It may be said that, even so, the court may remain less institutionally competent than the decision-maker in conducting the rationality review. This argument does not negate the possibility where institutional competence (or other factors, such as the extent of the allocative impact) can be relevant as a factor for judicial deference.[95] But judicial deference (building on Kavanagh's distinction earlier) is different from non-justiciability.[96] The foregoing analysis establishes that no clear-cut *binary* can be drawn between cases with an 'allocative impact' and 'discretionary allocative decision-making' – such that one may conclude that the former should not be *per se* non-justiciable, whilst the latter should (*per* the standard case) be categorically non-justiciable. In the lack of a good reason to sustain this analytical binary, both categories of cases should be justiciable – which therefore constitutes a baseline judicial role. But nothing said here precludes the possibility of judicial deference when rationality review is being applied: to defer in this more limited sense will *not* resurrect this analytical binary (to which this argument objects). Indeed, the need for deference has been accorded importance by the restrictive formula of the rationality review in *Wednesbury* – which enshrines an inherent

93 Wang (n 11 above) 635–636.

94 Lord Irvine of Lairg, 'Judges and decision makers: the theory and practice of Wednesbury review' [1996] Public Law 59, 60–61; Paul Craig, 'Reasonableness, proportionality and general grounds of judicial review: a response' (2021) 2 Keele Law Review 1, 3, 23; Craig (n 12 above) 24–25.

95 See eg Wang (n 85 above) 483–485.

96 Kavanagh (n 65 above) 244–245.

element of judicial restraint.[97] In light of this, it seems disproportionate to render judicial review unavailable altogether through a doctrine of non-justiciability.

It is, however, important to revert to the qualification hinted at earlier. The foregoing argument is predicated on the assumption that the court compels the reallocation of resources through adjudicating on healthcare resource allocation. But is this assumption sound? The literature has helpfully demarcated different possibilities upon which healthcare resource allocation is challenged. Wang and Newdick, for instance, have respectively recognised that such challenges may potentially result only in a 'procedural' remedy: the court will only 'quash the decision and remit the decision to the [decision-maker] for reconsideration'. This is to be distinguished from a 'substantive' remedy, whereby the court will make a court order for treatment to be provided.[98] This distinction is important: for the former remedy does not necessarily require the decision-maker to (upon reconsideration) reach a different decision: it may reach the same result (eg against the applicant), provided that it now meets all the legal requirements that it may have breached (when the decision was first struck down by the court).[99] In this case, it may plausibly be argued that the court has not compelled the decision-maker to reallocate resources – such that its actions will not amount to 'discretionary allocative decision-making'.

What may follow from this is that the argument in this section should be qualified: it may only apply insofar as the court applies a substantive remedy, but not a procedural remedy. But two points must be attached to this qualification, such that the effect of this should not be overplayed. First, the distinction between procedural and substantive remedies is not clear-cut. It has sometimes been said that the position in UK law is that only procedural – not substantive – remedies will be given.[100] But it has also been recognised that the line can often be blurred: for instance, the court may revert a decision for reconsideration – but has set such a high bar for the decision-maker that, in effect, it may well have prescribed a certain course of action.[101] Or short of this, it is possible to envisage a case that has generated so much media attention and political pressure that – after

97 Michael Taggart, 'Proportionality, deference, *Wednesbury*' [2008] New Zealand Law Review 423, 427–429; see also Elliott (n 74 above) 65-66; Paul Craig, 'Varying intensity of judicial review: a conceptual analysis' [2022] Public Law 442.

98 Wang (n 11 above) 641; Newdick (n 11 above) 112, 115–117.

99 See eg *R (SB) v NHS England* [2017] EWHC 2000 (Admin), [2018] PTSR 576 [105].

100 See eg Newdick (n 11 above) 111–112; Wang and Rumbold (n 41 above) 189; cf *R (S (A Child)) v NHS England* [2016] EWHC 1395 (Admin) [36].

101 Newdick (n 11 above) 112, 114–115.

the court has decided to quash the decision as being unlawfully made – the decision-maker may ultimately be pressured into granting the applicant funding, despite the procedural remedy given. In these cases where the procedural remedy has slipped into (what is in practice) a substantive remedy, the foregoing argument may still apply.

Second, it has been accepted throughout the literature that, if one is to be sceptical of the judicial role in healthcare resource allocation, it is cases with substantive remedies that are the most potentially problematic. Procedural remedies are by contrast more acceptable.[102] But this argument will mean that insofar as courts award substantive remedies – such that they compel the reallocation of resources – the doctrine of non-justiciability cannot justifiably apply, since no good reason exists to distinguish this type of case from cases with an allocative impact. Although this reasoning does not strictly cover cases when courts provide procedural remedies, it *must* follow from this that the doctrine of non-justiciability cannot equally apply here – for otherwise one would be accepting a *greater* degree of judicial restraint for cases awarding procedural remedies than in substantive remedies (which, as shown above, should be justiciable). Therefore, there must be a baseline judicial role – whether procedural or substantive remedies are being awarded.[103]

The case for justiciability: the *ultra vires* theory

The *ultra vires* theory entails that there should be, at least, a meaningful degree of judicial scrutiny over decisions concerning healthcare resource allocation. As is well-known, there has been a vibrant debate as to the constitutional foundation of judicial review.[104] One of the main contenders is the *ultra vires* theory. The theory maintains that decision-makers were conferred by Parliament only a limited jurisdiction. They must not be able to exceed that jurisdiction: doing so would mean they would be acting *ultra vires*.[105] This justifies the court's power to conduct judicial review: the courts are only enforcing

102 As shown in Wang (n 44 above) 715–721, 723–724; see also Newdick (n 11 above) 116–118.

103 This does not, of course, preclude discussion based on whether procedural or substantive remedies should be preferred: see Newdick (n 11 above) 116–118; Wang (n 44 above) 715–721, 723–724. The argument simply means that either way, the judicial role cannot be excluded for this reason.

104 See the discussion in Thomas Adams, '*Ultra vires* revisited' [2018] Public Law 31.

105 *Anisminic Ltd v Foreign Compensation Commission* [1969] 2 AC 147, 194–195; Dawn Oliver, 'Is the *ultra vires* rule the basis of judicial review?' in Christopher Forsyth (ed), *Judicial Review and the Constitution* (Hart 2000) 4.

the limits of jurisdiction,[106] thereby giving effect to what Parliament intends (ie the jurisdiction of the decision-makers must remain limited).[107] As Lord Sumption explained in *Privacy International*:

> If Parliament on the true construction of an enactment has created a tribunal of legally limited jurisdiction, then it must have intended that those limits should have effect in law. The only way in which a proposition can have effect in law, is for it to be recognised and applied by the courts.[108]

It follows, therefore, from the limited jurisdiction of a decision-maker's power that the decision-maker's discretion must not be unconstrained. A discretionary power that remains unchecked by judicial scrutiny means that it will not be limited.[109] This, however, will be the result of a doctrine of non-justiciability – whether it be a *de jure* or (as it is here) *de facto* doctrine of non-justiciability. No one suggests that health authorities have an unlimited jurisdiction. There can be no such suggestion, because the health authority *is* a public authority that has its limited powers derived from legislation.[110] There is thus a paradox: a decision-maker who has limited jurisdiction will be immune from meaningful judicial scrutiny. This contradicts the very essence of the *ultra vires* theory. As Farwell LJ observed, 'it is a contradiction in terms to create a tribunal with limited jurisdiction and unlimited power to determine such limit at its own will and pleasure—such a tribunal would be autocratic, not limited'.[111]

The critics may argue that a decision on healthcare resource allocation is different from other discretionary powers in that – *per* the standard case for non-justiciability – this is a matter that (a) has not been assigned by Parliament to courts[112] and (b) on which courts are not competent to adjudicate. But even if so – once an *ultra vires* analysis is applied – these concerns can no longer lead us to the conclusion that the decision should be non-justiciable. This is because

106 Christopher Forsyth, 'Of fig leaves and fairy tales: the *ultra vires* doctrine, the sovereignty of Parliament and judicial review' in Christopher Forsyth (ed), *Judicial Review and the Constitution* (Hart 2000) 30.

107 Paul Craig, '*Ultra vires* and the foundations of judicial review' in Christopher Forsyth (ed), *Judicial Review and the Constitution* (Hart 2000) 48; Christopher Forsyth and Mark Elliott, 'The legitimacy of judicial review' [2003] Public Law 286, 287.

108 *R (Privacy International) v Investigatory Powers Tribunal* [2019] UKSC 22, [2019] 2 WLR 1219 [210].

109 *Anisminic* (n 105 above) 194.

110 This has been the case throughout the history of the NHS: see Charles Webster, *The National Health Service: A Political History* 2nd edn (Oxford University Press 2002).

111 *R v Shoreditch Assessment Committee, ex p Morgan* [1910] 2 KB 859, 880.

112 See eg *Nottinghamshire County Council* (n 2 above) 247.

it remains that the jurisdiction of the decision-maker is limited. It may be that the court should allow more room for manoeuvre for a decision-maker. But there is no inherent qualification in the *ultra vires* theory that it does not apply in a socioeconomic context: it applies to all forms of limited powers conferred by Parliament on a decision-maker, *including* one to allocate resources. It is immaterial that courts are not as knowledgeable about healthcare resource allocation: the *ultra vires* analysis remains applicable, and the doctrine of non-justiciability will clearly contradict that. Nor is it material that this matter has been assigned by Parliament to government ministers: because the very idea of *ultra vires* is precisely premised on a primary duty being discharged by Government. The court's role has always been supervisory – with this premise in place.[113]

If we pause at this juncture, the foregoing analysis may face a formidable hurdle. We may conclude from the *ultra vires* theory that it is wrong for judicial review of healthcare resource allocation to be *entirely* unavailable – as is the case with the doctrine of non-justiciability. But this does not, at first sight, preclude the possibility that – as raised by Wang and Rumbold – we may exclude the *rationality review* of healthcare resource allocation, whilst maintaining the availability of procedural review.[114] In such a case, the health authority does not enjoy the unlimited power that Farwell LJ feared: the power of the health authority remains limited by the supervisory jurisdiction of the court, even though the full extent of judicial review may not be available. So, it seems that whilst this argument constitutes a positive case for *justiciability*, it does not go much further than that.

This conclusion appears intuitive, but there is more to the *ultra vires* theory that is of value here. The beauty of the *ultra vires* theory is that not only is the *existence* of judicial review justified by reference to legislative intent (as we have seen earlier), but that even the *controls* over discretionary powers (ie the grounds of review) were justified by reference to legislative intent.[115] This is so because – according to the leading proponents of the *ultra vires* theory, such as Allan, Forsyth and Elliott – Parliament does not stop at intending that the power of decision-makers must remain limited. Parliament also intends that discretionary powers must be exercised '*in accordance with the rule of law*'. The court's role is – in turn – to give specific content to the rule of

113 *Anisminic* (n 105 above) 194–195.

114 Wang and Rumbold (n 41 above) 186–189; see also Syrett (n 48 above) 480, 486; Wang (n 44 above) 723–724.

115 Craig (n 107 above) 49.

law, by developing the grounds of review in administrative law.[116] This way, the grounds of judicial review are all but reflective of legislative intent. For instance, Elliott suggested:

> *The rule of law*, which is a fundamental [principle] of the British constitution, *clearly favours the exercise of public power in a manner that is fair and rational*. It is entirely reasonable to assume that, in the absence of clear contrary enactment, *Parliament intends to legislate in conformity with the rule of law* … Thus Parliament, intending to legislate in conformity with the rule of law, is *taken only to grant such administrative power as is consistent with the requirement of that constitutional principle*. It is therefore taken to *withhold from decision-makers the power to act unfairly and unreasonably*, while recognising that the detailed requirements of fairness and rationality can most appropriately be determined by the courts through the forensic process.[117]

This passage connotes two propositions that are of great importance: (a) Parliament intends that decision-makers can only act in line with the rule of law; and (b) the rule of law requires both procedural fairness and rationality in decision-making. If *both* these propositions are accepted, it is clear that judicial review over decisions on healthcare resource allocation must at least include both procedural and rationality review – for otherwise we will risk defeating Parliament's intent to uphold the rule of law in the context of healthcare resource allocation. Or to put the same point in another way, courts must hold the decision-makers to the rule of law:[118] and this *requires* the existence of, *inter alia*, rationality review. This is, in itself, a direct and complete response to the critic's earlier point. The remaining analysis will perforce focus on whether these two propositions are indeed correct.

Proposition (a) is hardly disputable. As Elliott rightly explained, to suggest otherwise would be to suggest that Parliament is unconcerned with whether the rule of law is upheld. It is clearly more plausible to attribute to Parliament an intention that the rule of law should

116 Paul Craig, 'Competing models of judicial review' [1999] Public Law 428, 429–430 (emphasis added); Mark Elliott, 'The *ultra vires* doctrine in a constitutional setting: still the central principle of administrative law' in Christopher Forsyth (ed), *Judicial Review and the Constitution* (Hart 2000) 95, 98; Forsyth and Elliott (n 107 above) 287, 290; T R S Allan, 'The constitutional foundations of judicial review: conceptual conundrum or interpretive inquiry?' (2002) 61 Cambridge Law Journal 87, 104; T R S Allan, 'Constitutional dialogue and the justification of judicial review'(2003) 23 Oxford Journal of Legal Studies 563, 565.

117 Elliott (n 116 above) 95–96 (emphasis added).

118 A C L Davies, 'The administrative state and the fundamentals of public law' in Elizabeth Fisher, Jeff King and Alison Young (eds), *The Foundations and Future of Public Law* (Oxford University Press 2020) 257.

be upheld.[119] Allan agreed expressly with Elliott. To him, '[t]he preservation of the rule of law, as a basic protection against arbitrary power, is always an essential first premise': it is only right to reject the view that Parliament should be seen as 'neutral' about the manner in which discretionary power is exercised.[120] This is particularly true when viewed in light of the fact that the UK is a 'liberal democracy that preserves a basic separation of powers between the principal organs of government'; and with this constitutional context, it 'can be scarcely controversial' that Parliament will 'intend to honour the most fundamental requirements of the rule of law'.[121] And if any further proof is needed – as Lord Carnwath has rightly stressed in the recent landmark case of *Privacy International* – the rule of law has received express statutory recognition in section 1 of the Constitutional Reform Act 2005.[122] All of this provided solid proof for the correctness of proposition (a).

Let us then turn to proposition (b). Clearly, to deny the validity of the second proposition would be to deny that fair procedures and rationality are not 'dimensions of the rule of law'.[123] Since no one would seriously suggest that we should retain rationality review and remove procedural review – almost every academic in this field will gladly contradict this proposition[124] – the real controversy can really only be whether rationality in government is a dimension of the rule of law. This is what calls for some further thought here. Raz suggested that a 'commonly agreed' aim of the rule of law is to 'avoid *arbitrary government*'.[125] He later defines this conception of arbitrary government as follows: '[a]rbitrary government is the use of power that is indifferent to the proper reasons for which power should be used'.[126] There is thus an important relationship between the rule of law and the existence of reason. Endicott's work on this relationship is particularly instructive. He argues – like Raz does – that 'the rule of law is opposed to the arbitrary use of power':[127] '[a]rbitrary government is ... a departure

119 Elliott (n 116 above) 98 (emphasis added).
120 Allan, 'The constitutional foundations' (n 116 above) 104.
121 Allan, 'Constitutional dialogue' (n 116 above) 571–572.
122 *Privacy International* (n 108) [120].
123 Allan 'The constitutional foundations' (n 116 above) 99.
124 See eg Syrett, *Law, Legitimacy and the Rationing of Health Care* (n 4 above) 144–146, 231; Syrett, 'Healthcare resource allocation' (n 4 above) 117; Wang and Rumbold (n 41 above).
125 Joseph Raz, 'The law's own virtue' (2019) 39 Oxford Journal of Legal Studies 1, 5 (emphasis original).
126 Ibid.
127 Timothy Endicott, 'the reason of the law' (2003) 48 American Journal of Jurisprudence 83, 91.

from the rule of law, in favour of rule by the *mere will of rulers*'.[128] It is noteworthy that Endicott stressed that the rule of law is opposed to the 'mere' will of rulers: for to him, the defining feature of an arbitrary act is an act done *just because* 'the actor so wills' – and '*without* any (other) justification of *reason*'.[129]

It already follows from this that respect for the rule of law will naturally require a minimum degree of rationality in government decision-making. For, by combining the insights by Endicott and Raz, we can draw this conclusion: if we *are* to have the rule of law, we must avoid arbitrary government; and if we *are* to avoid arbitrary government, we must (by definition) ensure that government decision-making is rational. So, if judges are to safeguard the rule of law, this will call for the availability of rationality review.[130] This is why Endicott regards rationality review as an 'anti-arbitrariness doctrine': for through this doctrine, judges may demand that decision-making must be 'distinguishable from the mere arbitrary wills and private affections ... of the officials'. This doctrine is 'very closely allied to the rule of law because it gives the judges a way of standing against arbitrary decision making – and the rule of law, too, is opposed to the arbitrary use of power'.[131]

But there are two caveats to this analysis, which will ultimately qualify the baseline that this article is seeking to establish. First, this analysis only affirms the existence of rationality review; it does not preclude judicial deference when the rationality of decision-making is assessed. This is because – whilst courts must seek to prevent any arbitrary use of power – they should also 'do so in a way that gives the initial decision maker a leeway that corresponds to the reasons why the power was allocated to that person or institution'.[132] This is a concession that may be made, but this concession does not detract from the core thesis here: not only must decisions of healthcare resource allocation be *justiciable*, they must also be subject (at least) to both procedural *and* rationality review.

Second (which is related to the first caveat), this analysis does not prescribe the exact form and nature of rationality review. An important feature of the modified *ultra vires* theory is that courts are given much autonomy in defining the particulars of judicial review, given that

128 Timothy Endicott, 'Arbitrariness' (2014) 27 Canadian Journal of Law and Jurisprudence 49, 49 (emphasis added).

129 Endicott (n 127 above) 90 (emphasis added).

130 Ibid 91–92.

131 Timothy Endicott, 'Why proportionality is not a general ground of review' (2020) 1 Keele Law Review 1, 9–12, 23.

132 Endicott (n 24 above) 243.

these have not been developed by Parliament.[133] One implication of this is that various forms of judicial control – concerning the same ground of review – can be legitimated through the same analytical method.[134] Since the current baseline of the judicial role is developed by reference to the modified *ultra vires* theory, we can only infer through this line of reasoning that the rule of law *requires* (as we have seen) the *existence* of both procedural and rationality review. But the precise specificities of these grounds of review is another question and is not readily answered by reference to the modified *ultra vires* theory. One cannot conclude from the reasoning in this section whether courts should offer procedural or substantive remedies upon finding an illegal act,[135] or the precise form[136] and intensity of rationality review that should be adopted.[137]

That is: this thesis will not provide a ready answer to preferring one model of the judicial role in health litigation over another.[138] But we do know that there must be a *baseline*: that however valid concerns based on institutional competence and constitutional legitimacy are, there cannot be a doctrine of *de jure* or *de facto* non-justiciability; for this will surely contradict the requirements of the rule of law (embedded within the modified *ultra vires* theory) and the very point of an *ultra vires* analysis. Procedural and rationality review must at least be present as part of the baseline judicial role, and judicial deference (even if justifiably given – a possibility accepted in the first caveat) cannot amount to (either *de jure* or *de facto*) non-justiciability. This is the limited but significant contribution that this article makes.

CONCLUSION

To conclude, three propositions have been made. First, concerns in favour of judicial restraint – even if valid – do not justify a doctrine of non-justiciability. This doctrine is inflexible and is perforce unjustified. Second, we can all agree that cases with an allocative impact are justiciable. Since we cannot sensibly distinguish decisions

133 Forsyth and Elliott (n 107 above) 287.

134 This feature of the model has indeed been turned into a critique of it: see Craig (n 91 above) [1-012], [1-016].

135 See the discussion in Newdick (n 11 above) 116–118; Wang (n 44 above) 715–721, 723–724.

136 See eg the various possibilities of understanding rationality review outlined in Yossi Nehushtan, 'The true meaning of rationality as a distinct ground of judicial review in United Kingdom public law' (2020) 53 Israel Law Review 135; see also Daly (n 92 above) 242–247; Hasan Dindjer, 'What makes an administrative decision unreasonable?' (2021) 84 Modern Law Review 265.

137 Craig (n 12 above) 25–26.

138 See the models as outlined in Newdick (n 11 above); Wang (n 44 above).

concerning healthcare resource allocation from such cases, the doctrine of non-justiciability cannot be sustained. We can tell from the first two propositions that the standard case for non-justiciability is flawed. Third, the *ultra vires* theory entails that not only must decisions concerning healthcare resource allocation be justiciable: they should also be subject to both procedural and rationality review in the UK courts, by reference to the requirements of the rule of law. These propositions together establish a baseline judicial role over healthcare resource allocation in UK law and ultimately *justify* the move from non-justiciability (as posited by the first batch of case law) in the current UK jurisprudence.

There are two messages that underlie this article that may be useful for broader purposes. First, it is important to remember – for academics in medical law and public law alike – that the NHS *is* a public authority that derives its power from legislation and is thus subject to the rule of law.[139] This means that – despite all the good things that may be said of the NHS – external control must be imposed to ensure that it measures up to what the rule of law requires.[140] The departure from the doctrine of non-justiciability over decisions concerning healthcare resource allocation is one facet of this: but this overarching message should be borne in mind in a much broader range of contexts. Second, this article clearly does not offer a comprehensive theory of the judicial role in healthcare resource allocation in UK law. This article only offers a baseline judicial role that must be maintained, concerns of institutional competence and constitutional legitimacy notwithstanding. But it does not mean this article is entirely irrelevant to the development of such a comprehensive theory: for instance, the discussions here on the need for flexibility in judicial restraint may be relevant to ascertaining the appropriate intensity of review; the discussion here also tells us that such a comprehensive theory *cannot* violate the baseline established here.

139 Davies (n 118 above) 257.
140 Ibid.

Northern Ireland
Legal Quarterly
Winter Vol. 73 No. 4 (2022) 618–649
Article DOI: 10.53386/nilq.v73i4.950

Rethinking dispute resolution mechanisms for Islamic finance: understanding litigation and arbitration in context

Abdul Karim Aldohni

Newcastle University
Correspondence email: a.k.aldohni@newcastle.ac.uk

ABSTRACT

While there seems to be a growing appetite for Islamic finance products at a global level, the parties using these products do not seem to pay enough attention to how best they can resolve any disputes arising from these agreements. It is a shortfall that undermines the Islamic compliance aspect of these transactions and jeopardises their unique Islamic characteristic. This article considers ways in which English litigation can be used as an optimal mechanism to resolve Islamic finance disputes. The article particularly analyses the incorporation of international Sharia Standards in Islamic finance agreements as a way to overcome the disadvantages of ligation highlighted by a large body of case law in this context. It then argues that, while arbitration might seem on the face of it a more appropriate mechanism, it is riddled with complexities and disadvantages.

Keywords: Islamic finance disputes; English court litigation; arbitration; international standards incorporation.

INTRODUCTION

The Islamic finance industry has enjoyed significant growth, particularly in the last two decades. It is estimated that the industry is currently worth $2.2 trillion with an expected continuous growth rate in 2022/2023 of about 10 per cent. Despite the double shock from the Covid pandemic and the drop in oil price, the industry grew rapidly in 2020 albeit at a slower rate compared to 2019.[1]

The serious business credentials of the industry have allowed it to become a global rather than regional industry and to attract international investors from the entire globe. Further, it has opened up the door of some of the major financial centres in the Western world. An example in point is the United Kingdom (UK), where the Government has long taken a special interest in developing its Islamic

1 S&P Global Ratings, Islamic Finance Outlook (2022 edition) and S&P Global Ratings, Islamic Finance 2022–2023: Same Constraints, New Opportunities.

financial sector. The position of London as a world-leading financial centre has attracted Islamic financial institutions and also provided the City of London with a variety of new and innovative Islamic financial products. This mutual interest between the UK and the Islamic finance sector has been manifested in the offering of Islamic financial products by a number of high-street conventional banks, which have paved the way for the UK financial market to host a number of fully fledged Islamic banks. Further, in June 2014, the UK Government was the first one outside of the Islamic world to issue sovereignty Sukuk al-ijara worth £200 million, which matured on 22 July 2019. Given its success, the UK Government followed it in 2021 by a second sovereignty Sukuk al-ijara offering worth £500 million, which matures on 22 July 2026.[2]

This transformation from a regional to a global industry has come with some serious legal challenges. The regional familiarity and, to an extent, acceptance of the industry's legal foundation, namely Islamic law, is no longer a given. On the contrary, Islamic law, 'Sharia', is not a recognised source of law in Western jurisdictions. This, in truth, primarily stems from the nature of Islamic law in its present form, as an abstract concept. Islamic law, conceptually, is widely understood and accepted as a divine law founded in the religion of Islam. Practically, however, it lacks the systemisation and standardisation that creates structure and certainty and can be only offered by the sovereignty of a state.[3] In fact, even in a jurisdiction such as Saudi Arabia that claims to be Sharia-based, the state has not developed the essential foundations and processes to achieve the required structure and certainty for Islamic law.[4] Therefore, contemporary reference to 'Islamic law' entails the reference to a collection of principles and rules found in the Quran and Prophetic Sunnah,[5] on the one hand, and the broad scholarly work of Muslim jurists to interpret and apply these principles and rules on

2 HM Treasury, 'UK bolsters Islamic finance offering with second Sukuk' (25 March 2021).

3 Wael B Hallaq, *The Impossible State* (Columbia University Press 2013) 30–31.

4 Hossein Esmaeili, 'On a slow boat towards the rule of law: the nature of law in the Saudi Arabia legal system' (2009) 26(1) Arizona Journal of International and Comparative Law 1–47.

5 The Quran is a direct divine revelation which is believed to be the words of God '*Allah*' that were revealed to his last messenger Prophet Muhammad. It is the highest and most authenticated source of Islamic law as it was recorded in writing during the Prophet Muhammad's lifetime. The Prophetic Sunnah is the second source of divine revelation as mandated in the Quran [53:3–4]: 'Nor does he say [aught] of [his own] Desire. It is no less than revelation sent down to him.' It encompasses all the Prophetic statements and actions that were narrated by his companions, later collected, and recorded in writing by their followers.

the other.[6] This engenders variation and uncertainty in the absence of divine revelation, as there is no qualified authority to take over the divine legislative power, or to act as an authoritative interpreter in the way a final court can.

In the context of Islamic finance, there are two dimensions to this identified legal challenge; the first concerns the operations of Islamic finance in these Western global financial centres, which is primarily a regulatory challenge that the article does not address. The second concerns settling the disputes that arise from Islamic finance operations, which is the focus of this article.

As mentioned earlier, one important aspect of becoming a global industry is that Islamic finance has become a means to facilitate international investments. Parties from different jurisdictions can now agree to use an Islamic finance product to facilitate a commercial or financial transaction, which is not necessarily executed in their jurisdictions. This mandates the inclusion of a governing law clause, according to which the parties mutually choose a jurisdiction to determine their rights and obligations under the contract. In this respect, the wide interpretation of the freedom of contract principle under English law and the trustworthiness of the English judiciary have induced parties to international Islamic finance transactions to elect the English law and its court system as their choice to settle their disputes.

As this article is set to rethink the best mechanism to settle Islamic finance disputes, it is divided as follows. The next part examines the use of litigation before the English court in this respect. Drawing on a host of case law, it narrows down this challenge to two issues. First the classification of Islamic law under the English legal system and, second, the technicality of proving Islamic law before the court. Accordingly, it proposes a solution for the parties to an Islamic finance agreement to consider in advance if litigation is their preferable route to resolution while they remain committed to the Sharia integrity of their transaction. In other words, the focus is *how* best to litigate rather than *why* not litigate an Islamic finance dispute before the English court. The article then examines whether arbitration could be a more optimal alternative to litigation in the context of Islamic finance disputes. It argues that although arbitration might, at first, seem a more straightforward solution, it has its many challenges that make it far from perfect. It

6 This represents the human endeavour (*ijtihad*) to understand the divine textual sources and apply their rulings to ever-evolving circumstance by using human reasoning and logic. Although this process is governed by the rules of Islamic jurisprudence (*usul alfiqh*), it remains highly susceptive to subjectivity associated with the personal input of each and every jurist or scholar. See A K Aldohni, 'A compatibility analysis of Islamic financial disputes: English private international law and Islamic law' (2019) 14(1) Journal of Comparative Law 219–221.

is important to note that the analysis of arbitration only concerns an arbitral process where England is the seat. The final part concludes the discussion by bringing together the key arguments made in this article.

LITIGATION: FOREIGN LAWS, ISLAMIC LAW AND THE ENGLISH COURT

The English legal system has developed over the centuries its unique private international law framework. This set of rules is shaped by a historic narrative founded in the decisions of the common law courts, the adversarial character of the English legal system and international treaties. Taking into account these unique features and reflecting on a large body of case law, the below discussion demonstrates how best to litigate Islamic finance disputes before the English courts.

For historic reasons concerning the evolution of common law,[7] English law does not assume that all laws are equal partners 'in the community of law of nations'.[8] The English court's knowledge of all laws only extends to English law and excludes foreign laws.[9] Therefore, the English court will not introduce the rules of a foreign law *ex officio*.[10] This is not to suggest that the English court is not equipped to apply a foreign law, English private international law accommodates for this once two requirements are satisfied: first, that the foreign law is in itself applicable and, second, that its contents have been proved to the satisfaction of the court.

Applicable foreign law

English private international law refers to certain cases where a foreign law could be potentially applicable, provided the parties plead it: for example, tortious disputes concerning a personal injury that took place in a foreign jurisdiction (*lex delicti*); proprietary disputes concerning a property located in a foreign jurisdiction (*lex situs*); and contractual matters concerning the choice of a foreign law as the contract governing law, under which *Shamil Bank of Bahrain v Beximco Pharmaceuticals*

7 Fentiman suggests that historically common law courts had jurisdiction over domestic disputes where they applied what later became known as English common law, while any other cases with foreign elements fell within the jurisdiction of admiralty courts. This arrangement later changed as a result of the eventual dominance of the common law, English law and courts for that matter would – in principle – only recognise and apply the rules of English law to all disputes. See R Fentiman, 'Foreign law in English courts' (1992) 108(1) Law Quarterly Review 142–156.

8 Ibid 143.

9 R Fentiman, *Foreign Law in English Courts: Pleading, Proof and Choice of Law* (Oxford University Press 1998) 5.

10 Ibid 68.

Ltd and Others[11] and *Dubai Islamic Bank PJSC v PSI Energy Holding Company BSC*[12] fall, at least on the face of it. It is important to note that including a conclusive choice of a foreign law in the terms to govern the agreement still does not guarantee its application by the English court. As seen in *Aluminium Industrie Vaassen B V v Romalpa Aluminium Ltd*,[13] unless either of the parties were to invoke the choice of law clause, it would have no effect and English law will apply.[14]

In *Shamil Bank of Bahrain v Beximco Pharmaceuticals Ltd and Others*,[15] the parties agreed to use a *Murabaha* (mark-up) agreement as a facility for trade finance. Under this type of agreement, the Islamic financial institution buys the required goods desired by the client. Upon obtaining the goods, the bank resells the goods to the client with an added margin of profit to the original purchase price. Given that the Islamic financial institution is only paid once the goods are supplied to the client, the Islamic financial institution bears the risk of delivery failure. In theory, this genuine form of risk-sharing is what justifies the added margin of profit from an Islamic law perspective. Accordingly, Shamil Bank under a *Murabaha* agreement with the first defendant undertook to buy goods and to resell them to the first defendant for an added margin of profit. The second defendant was appointed by the Bank as its agent 'for the purchase of the goods'.[16] As for the governing law clause, the parties expressed that all legal issues 'arising out of or in connection to the agreement'[17] should be 'subject to the principles of Glorious Sharia, this agreement shall be governed by and constructed in accordance with the laws of England'.[18]

The dispute was brought before the English court because of a default in payment by the defendants, who argued that the overdue payment and any agreed compensation were not enforceable, citing the governing law clause.[19] The defendants argued that both English and Sharia laws should apply, while English law would sanction such a

11 [2003] EWHC 2118 (Comm).
12 [2013] EWHC 3186 (Comm).
13 [1976] 1 WLR 676.
14 Despite having clause 30 in the disputed agreement, which subjected the conditions of the agreement to Dutch law and gave the Amsterdam court an exclusive jurisdiction, the English court decided that English law is the applicable law. Neither the plaintiffs, a Dutch company, nor the defendants, a British company, pleaded Dutch law as the governing law of the agreement by invoking the choice of law clause *Aluminium Industrie Vaassen B V v Romalpa Aluminium Ltd* [1976] 1 WLR 676, 684. See also Fentiman (n 7 above) 149.
15 [2003] EWHC 2118 (Comm).
16 Ibid para 4.
17 Ibid cited in para 5.
18 Ibid cited in para 5.
19 Ibid cited in para 15.

payment, Sharia would prohibit it, as it constitutes the forbidden '*riba*' (interest).

While the defendants were quick to plead the application of 'glorious Sharia', which is procedurally essential, there were substantive failures in the choice of law clause that made it ineffective. First, the freedom to choose a governing law must be 'affirmatively used',[20] therefore the choice clause must be structured clearly and conclusively. The first instance court found that was not the case. Mr Justice Morison stated that 'it cannot have been the intention of the parties that it would ask this secular court to determine principles of law derived from religious writing on matters of great controversy'.[21] He found it highly improbable that the intention of the parties was to ask the English court 'to determine difficult questions of the Sharia principles',[22] which include 'conflicting pronouncements' and many of the commercial issues which are still quite debatable.[23] The fact that the court had to guess the intention demonstrates the lack of an affirmative choice of Islamic law or 'glorious Sharia'.

This leads to the second issue that is the legal classification of 'Islamic law' or 'Sharia' in light of the meaning of a valid choice of law. It has been long established under common law,[24] then the Rome Convention[25] and now the Rome I Regulation,[26] that only a national system of law, namely the law of a country, could be a valid choice

20 This requirement has long been under the Rome Convention on the Law Applicable to Contractual Obligations (Rome Convention), which came into force in the UK after its implementation by the Contract (Applicable Law) Act 1990, art 3(1), and then later in the Law Applicable to Contractual Obligations (Rome I) Regulation (EC) No 593/2008 (Rome I Regulation), art 3(1). For more discussion, see also Adrian Briggs, *The Conflicts of Law* 4th edn (Oxford University Press 2019) 214.

21 *Shamil* (n 11 above) cited in para 35.

22 Ibid cited in para 24.

23 Dr Lau's (expert witness) opinion on Sharia law: ibid para 24.

24 In *Amin Rasheed Shipping Corporation v Kuwait Insurance Co* [1984] AC 50, 61, Lord Diplock – while quoting Lord Simonds' 'pithy definition' of the proper law of contract in *John Lavington Bonython and Others v Commonwealth Australia* [1952] AC 201 – stressed that under English conflict rules the 'proper law' of contract 'is the substantive law of the country which the parties have chosen that the courts of that country might themselves apply if the matter were litigated before them'.

25 Rome Convention, art 1(1) (emphasis added): 'The rules of this Convention shall apply to contractual obligations in any situation involving a choice between the *laws of different countries*.'

26 Rome I Regulation, art 3(3) (emphasis added): 'the *country whose law has been chosen*'.

of governing law in a contract.[27] This automatically disqualifies the choice of Islamic law or any other religious law for that matter, such as 'Jewish law' in *Halpern v Halpern*,[28] as the court does not have the power to give effect to this choice. The court in the *Shamil Bank* case reached the same conclusion, citing the applicable law at the time – the Rome Convention.[29]

Dubai Islamic Bank PJSC v PSI Energy Holding Company BSC[30] has a strong sense of *déjà vu* about it. The case concerned a debit-restructuring agreement the parties reached in relation to the outstanding principal amount, which the bank gave in the course of a legitimate Sharia-compliant agency agreement, and the agreed profits. The governing law clause in the disputed debt-restructuring agreement chose English law to govern the agreement 'save in so far as inconsistent with the principles of Sharia law'.[31] The defendants argued that making the agreed payments – the principal amount and profits – would breach Sharia, and, given the governing law clause, the payment under English law was no longer enforceable. The court decided that such a 'proviso is of no effect', citing the decisions of the *Shamil bank* and *Halpern* cases, and subjected the agreement to the exclusive jurisdiction of the English court.[32]

Satisfactory proof

Although Islamic law does not qualify as a legitimate choice of governing law in a contract, in theory, nothing stops the parties to an agreement from choosing the law of a country – a national system of law – that incorporates some elements of Islamic law. The reality, however, is far from simple as such incorporation tends to be limited in its coverage and does not guarantee the Sharia compliance of the entire laws of that state. For instance, the United Arab Emirates (UAE) Civil Code, which is influenced by several civil codes in the region that are primarily based on the French Civil Code,[33] incorporates elements of

27 There could be one governing law of a particular country to ensure certainty, for detailed scholarly commentary on the case, see Jason Chuah, 'Islamic principles governing international trade financing instruments: a study of the Morabaha in English law' (2006) 27(1) Northwestern Journal of International Law and Business 137–170, 144.

28 [2008] QB 195.

29 The court stated that: 'Article 1(1) of the Rome Convention makes it clear that the reference to parties' choice of the law to govern a contract is a reference to the law of a country': *Shamil* (n 11 above) para 27.

30 [2013] EWHC 3186 (Comm).

31 Ibid para 11.

32 Ibid para 11.

33 *Glencore International AG v Metro Trading International Inc* [2001] CLC 1732, 1751.

Islamic law concerning loan (*qardh*) contracts,[34] transfer of ownership in bailment[35] and misappropriation.[36] On the other hand, the UAE Commercial Code allows for the payment of interest on commercial loans and delay interest on commercial loans and commercial obligations fixed in a sum of money.[37] Therefore, the 'law of the UAE' could be a valid choice of governing law yet it would not guarantee that 'Islamic law' exclusively governs the agreement. Further, if the law of such a country were made as an express choice of the governing law, any modifications that changed the law to non-Sharia compliant would still bind the parties.[38]

More importantly, the English court will treat the foreign law as 'a peculiar kind of fact'[39] that has to be proved to the court's satisfaction.[40] This is something that the court has dealt with on numerous occasions where the proprietary[41] or tortious[42] disputes connected the case, according to English private international law, to foreign jurisdictions that included elements of Islamic law.

In an adversarial legal system, such as the English legal system, the parties to a dispute will call their witness to give statements[43] concerning the application of the foreign law while the court is not actively involved in this fact-finding process. The court will assess the evidence provided and decide, accordingly, how the foreign law applies in the context of the case. It is the parties who appoint their expert witness rather than the court, therefore, the expert evidence is likely to conflict as each expert advocates a more favourable position of their

34 Arts 992–993.

35 Art 975.

36 Art 1326.

37 Commercial Transactions Code, arts 77, 78, 88.

38 Lord Collins of Mapesbury and J Harris (eds), *Dicey, Morris and Collins on the Conflict of Laws* 15th edn (mainwork) and 5th supp (Sweet& Maxwell 2018) vol 2 (32-058).

39 *Parkasho v Singh* [1968] P 233, 250. Sir Jocelyn made this remark while explaining that the appellate court will still have the power to interfere with the finding of the trial court concerning the question of foreign law despite being classified as a fact. The appellate court can assess whether the evidence justifies the trial court's conclusion regarding the question of foreign law, which does not normally extend to the other relevant facts in that case, and see also Fentiman (n 7 above) 145.

40 *Guaranty Trust Company of New York v Hanny & Co* [1918] 2 KB 623 and *Ertel Bieber & Co v Rio Tinto Co Ltd Dynamit AG (Vormals Alfred Nobel Co) v Rio Tinto Co Ltd Vereingte Koenigs v Rio Tinto Co Ltd* [1918] AC 260.

41 For example, *Glencore* (n 33 above).

42 For example, *Harley v Smith* [2010] CP Rep 33 and *Abdel Hadi Abdallah Al Qahtani & Sons Beverage Industry Company v Andrew Antliff* [2010] EWHC 1735 (comm), Official Transcript.

43 A A Gillespie and S Wear, *The English Legal System* 5th edn (Oxford University Press 2015) 14–15.

party.[44] This could prove particularly problematic in the context of Islamic finance disputes.

On the one hand, the issue of Sharia compliance is central to the dispute; on the other hand, the judge's use of the evidenced Islamic law is solely based on the quality of the expert's evidence statement and performance in court. These are two fundamentally different things and the high quality of the latter does not necessarily guarantee achieving the former. For instance, *Abdel Hadi Abdallah Al Qahtani & Sons Beverage Industry Co v Andrew Antliff*[45] concerned the meaning of bribery and duty to declare conflict of interests in Sharia under Saudi law. While the judge acknowledged the expertise of the claimant's expert witness as a Sharia scholar and Saudi law practitioner, he questioned his abilities to provide consistent explanations.[46] It can be suggested that such an impression by the court has its impact on the extent to which the witness statement is factored in the judge's ruling. Therefore, appointing the more convincing expert witness, who is probably the more expensive, could be a decisive factor in these cases. Further, it can be argued that fluency in the English language and the ability to explain the complicated Sharia points in question in a legal language that the court is familiar with are qualities that contribute immensely to the strength of the statement, which does not always correlate with its Sharia rigour. In other words, it is not always the case that the expert who has the Sharia training and the language skills to interrogate the vast Islamic jurisprudence literature available in Arabic also has the English language skills to convey this knowledge to the court.

An English judge cannot decide by himself or herself what the foreign law means without relying on the proof provided by the parties. In *Harley v Smith*,[47] the case concerned a tortious claim brought by three British former employees (professional divers) who were injured in Saudi territorial waters. The court considered Islamic law to decide the meaning of 'work relation' and whether the claim was time barred. The first instance court concluded that the time limitation should be interpreted in line 'with the Sharia principles of there being no limitation period (or at least none as short as one [year]) in relation to ordinary personal injury claims'.[48] The Appeal Court particularly criticised this finding as it found that there was no evidence presented to the court to support the judge's interpretation of the 'work relation'

44 T C Hartley, 'Pleading and proof of foreign law: the major European systems compared'(1996) 45(2) International and Comparative Law Quarterly 274.

45 *Abdel Hadi Abdallah Al Qahtani & Sons* (n 42 above).

46 Ibid para 29.

47 [2010] C Rep 33

48 *Harley v Smith* [2009] PIQR P11, para 82.

under Sharia. Therefore, in the absence of the required proof for this fact (ie a Sharia or Islamic law extended meaning of work relation) the judge 'decided for himself what Sharia law would require' and 'went beyond what he could properly do' in construing 'foreign legislation by applying principles of interpretation which had not been established by evidence'.[49]

Therefore, it is inherent in an adversarial legal system that the court's view of the substance of the foreign law is not established in complete isolation from the parties' influence, albeit through the legitimate means of expert witnesses. This situation can only be avoided where the judge is primarily tasked with establishing the substance of the applicable foreign law. Take for example the German legal system. Section 293 of the Code of Civil Procedure (*Zivilprozessordnung* – ZPO) requires the court to determine the substance of the foreign law using the court's own research. This may include contacting 'the competent authority in the foreign state concerned' and obtaining a legal opinion from an expert.[50] However, the use of an expert's opinion in this context is quite different from that under the English legal system. The court objectively appoints the expert based on their knowledge and practical experience of the foreign law, which prevents the parties from shopping for a favourable opinion. Accordingly, in an Islamic finance dispute this prevents a party who can afford a favourable opinion with a convincing quality – not necessarily matched with its Sharia rigour – from succeeding.

Based on the above, there are some key observations to make. First, it is well established that parties to an Islamic finance dispute are not legally entitled to ask the English court to apply a non-national system of law. Therefore, it can be suggested that the superficial use of terms such as 'Islam law' or 'Sharia' in the governing law clause is, in a way, an affirmative choice by the parties not to apply Sharia. Second, even the choice of a national system of law that includes elements of Islamic law brings a host of concerns that may undermine the Sharia compliance goal. Accordingly, the optimal solution to this problem rests on two factors: on the one hand is ensuring the certainty of the Sharia-based rules that the English court is authorised by private international law to apply; on the other is minimising the influence that the parties can exert, albeit legitimately, over the court's understating of the substance of these rules.

In this regard, it is argued that this still can be achieved in litigation before the English court through incorporating Sharia-based principles in the terms of the contract, which is not as simple as it may seem. In

49 *Harley v Smith* (n 47 above) para 50.
50 European e-Judicial Portal, 'Germany'.

order for incorporation to succeed, there are procedural and substantive matters that require careful consideration and adherence.

In English contract law, parties to a contract may elect to incorporate certain terms into their agreement as long as a number of procedural hurdles have been overcome. First, is ensuring that a notice of the term(s) in question is given before or at the time of concluding the contract: second, ensuring that the document – intended to have contractual effect – contains the incorporated term(s); third, that reasonable steps were taken to bring the term(s) to the attention of the parties;[51] and, finally, ensuring that the incorporation takes place in the actual terms (ie the operative parts) of the contract.

Islamic Investment Company of the Gulf v Symphony Gems NV & Ors[52] could be a good example of the failure to comply with the procedural requirements of incorporation. The case concerned a *Murabaha* (mark-up) agreement, in which the parties agreed, in a clear choice of law clause, that: 'this Agreement and each purchase agreement shall be governed by, and shall be constructed with, English law'. They also agreed that: 'the courts of England shall have exclusive jurisdiction to hear and determine any suit'.[53] Nevertheless, one of the recitals to the agreement stated: 'The Purchaser [Symphony Gems] wishes to deal with the Seller [IICG] for the purpose of purchasing supplies under this Agreement in accordance with the Islamic Sharia.'[54]

The case was brought to court with regard to the amount of the balance due to IICG. The defendant (Symphony Gems, ie the purchaser) argued that the payment default was due to an alleged delivery failure by the supplier, a risk that should be borne by the claimant (IIGS, ie the seller) according to Islamic law. The reference to Islamic law in this agreement was not made in the terms, rather it was in one of the recitals that could only play a role if 'the operative part' of the agreement was ambiguous.[55] Hence, it was unenforceable given that the operative parts were clear in nominating the English law as the governing law and the English court as the one with exclusive jurisdiction to settle any related disputes. Therefore, the 'Islamic Sharia' recital could not control the clear governing law term in this case.[56] The judge stated: 'it is a contract governed by English law. I must simply construe it according to its terms as an English law contract.'[57]

51 E Mckendrick, *Contract Law* 13th edn (Palgrave Macmillan 2019) 172.
52 [2002] 2 WLUK 313.
53 Cll 25–26 of the contract cited in ibid.
54 Ibid.
55 H G Beale (ed), *Chitty on Contracts* 34th edn (Sweet & Maxwell 2021) vol 1, 1149.
56 Ibid.
57 *Islamic Investment Company* (n 52 above).

This leads to the substantive requirements concerning the terms that incorporate Sharia principles in an Islamic finance agreement, which should be considered in light of English private international law. Recital 13 of the Rome I Regulation states: 'this regulation does not preclude parties from incorporating by reference into their contract a non-State body of law or an international convention'.[58] In principle, this suggests that parties to an Islamic finance agreement can elect to incorporate elements of Islamic law into the terms of their contract provided they fulfilled the earlier discussed procedural requirements. However, in order for such incorporation to take effect from a private international law perspective it must identify specific 'black letter' provisions.[59] Therefore, in a case such as *Symphony Gems*[60] the reference to 'Islamic Sharia', even if it were made in the terms of the agreement, would have been ineffective. Similarly to terms such as 'glorious Sharia' and 'Islamic law', it still lacks the certainty of sufficiently identified 'black letter' provisions.[61]

It is argued, therefore, that the Sharia Standards of the Accounting and Auditing Organisation for Islamic Financial Institutions (AAOIFI) should be used for effective contractual incorporation.[62] On the one hand, AAOIFI is the leading international not-for-profit organisation primarily responsible for developing and issuing standards for the global Islamic finance industry. It has institutional members from over 45 countries including central banks and financial regulators.[63] On the other hand, AAOIFI Sharia Standards represent a set of rules that clearly articulates the underpinning Sharia principles of a large array of Islamic finance agreements and products. These international standards have been either mandated or adopted by a number of regulatory authorities around the world.[64]

58 For detailed analysis of recital 13 and its legislative history, see Chuah (n 27 above) 195–196.

59 Collins and Harris (n 38 above) vol 2 (32-056)–(32-058); see also Chuah (n 27 above) 150–151.

60 Islamic Investment Company (n 52 above).

61 For detailed discussion on legal risk and uncertainties associated with Islamic law, please see Andrew White and Chen Mee King, 'Legal risk in Islamic finance' in Simon Archer and Rifaat Ahmed Abdel Karim (eds), *Islamic Finance: The New Regulatory Challenge* 2nd edn (Wiley 2013) 226–228.

62 Rupert Reed, 'The application of Islamic finance principles under English and DIFC Law' (2014) (Oct) Butterworths Journal of International Banking and Financial Law 574–575.

63 AAOIFI, 'About AAOIFI'.

64 For example, from 1 September 2018 the UAE Central Bank required that all fully fledged Islamic banks, Islamic windows of conventional banks, and finance companies offering Sharia compliant products and services in UAE must comply with the AAOIFI Standards. AAOIFI, 'AAOIFI welcomes UAE's adoption of its Standards'.

The incorporation of these Sharia Standards in the terms of an Islamic finance agreement would certainly fulfil the earlier identified requirement for effective incorporation. In addition to having the required international recognition and authority, these standards have 'black letter' provisions that are published by AAOIFI in clear and understandable English. For example, AAOIFI's Sharia Standard No 8 provides clear and precise Sharia provisions regarding a number of questionable issues in the practice of a *Murabaha* (mark-up) agreement, a widely used agreement by Islamic financial institutions. Some of these issues were disputed in both the *Shamil Bank* and *Symphony Gems* cases in the context of Sharia compliance.[65] This includes the requirement that the Islamic financial institution 'must assume the risk of the item it intends to sell',[66] which encompasses any risks associated with the delivery and possession of the item.[67] In addition, it prohibits the seller from subsequently demanding an extra payment in consideration for delay in payment.[68] Unfortunately, the terms of the agreements in the *Shamil Bank* and *Symphony Gems* cases[69] were in clear contradiction to the provisions of AAOIFI Sharia Standard No 8 and the superficial reference to 'glorious Sharia' and 'Islamic law' had no effect.

Although the use of incorporation of international standards has not been tested in the context of Islamic finance disputes litigation before the English court, there are many examples of incorporated international commercial standards, especially in the context of arbitration, in which the English court upheld its implementation. For instance, in *Econet Satellite Services Ltd v Vee Networks Ltd*[70] the parties signed a number of agreements, including 'the Main Contract', concerning the supply of satellite equipment and technology. Later, they signed a further agreement, the Voice Traffic Termination Rate Agreement (VTTRA), in which they included a governing law clause choosing 'the substantive internal laws of the United Kingdom applicable to contracts executed and to be ... interpreted in accordance with the UNIDROIT Principles of International Commercial Contracts'.[71] More importantly, the VTTRA included an arbitration clause that incorporated the United Nations Commission on International Trade Law (UNCITRAL) Arbitration Rules.[72]

65 *Shamil* (n 11 above) and *Islamic Investment Company* (n 52 above).
66 AAOIFI, Sharia Standard No 8 3/1/1.
67 Ibid 2/5/2.
68 Ibid 4/8.
69 *Shamil* (n 11 above) and *Islamic Investment Company* (n 52 above).
70 [2006] 2 CLC 488.
71 S 15(1) cited in ibid para 5.
72 S 16(1) cited in Ibid.

The VTTRA was disputed in relation to unpaid invoices, therefore, an arbitration began in which the party owing the payment admitted the sum claimed but by a way of defence made a counter-claim of set-off arising out of 'the Main Contract' and not the VTTRA. The arbitrators decided that article 19(3) of the UNCITRAL Arbitration Rules[73] incorporated in the VTTRA would only allow a set-off based on a counter-claim arising out of the *same* contract in question: that is only out of VTTRA. This arbitral award was challenged before the English court on a point of law under section 69 of the Arbitration Act 1996 (discussed in detail in the next section) given that the chosen governing law was the laws of the UK. The court confirmed that 'the meaning and effect of Article 19 (3) is clear',[74] therefore, the arbitral tribunal was right not to allow a set-off counter-claim. In the court's opinion, the parties agreed to the effect of article 19(3) by signing up to the incorporation clause in the VTTRA.[75]

It is argued, therefore, that the incorporation of AAOIFI Sharia Standards provides the clarity and certainty that allow the court to deal effectively with highly contested Sharia matters and it demonstrates the parties' true commitment to Sharia compliance. In addition, it is the judge who interprets these terms as terms of an English contract. And given that they are published in English in a precise, 'black letter', form, it is suggested that the Sharia compliance of the agreement will be certainly far more observed by the incorporation of these standards. This is because the earlier identified challenges associated with the use of expert witnesses in the context of applying a foreign law will no longer be a matter of concern.

ARBITRATION AND ISLAMIC FINANCE DISPUTES

It is well established that arbitration provides an effective alternative mechanism to resolve commercial and financial disputes. In this respect, the concept of arbitration is deeply rooted in the primary sources of Islamic law, the Quran and the Sunnah, as a recommended form to resolve disputes in general without specific reference to commercial disputes. For example, the breach of the Quranic prohibition of hunting during the pilgrimage period would result in a fine which the Quran requires to be estimated by a just arbitrator.[76] In the context

73 UNCITRAL Arbitration Rules, art19(3): '3. In his statement of defence, or at a later stage in the arbitral proceedings if the arbitral tribunal decides that the delay was justified under the circumstances, the respondent may make a counter-claim arising out of the same contract or rely on a claim arising out of the same contract for the purpose of a set-off.'

74 *Econet* (n 70 above) para 21.

75 Ibid para 21.

76 Quran, verse [5:95].

of domestic disputes, the Quran also requests the husband and wife to resort to arbitration.[77] Further, the Prophetic Sunnah includes a number of references to the Prophet's use of arbitration in numerous warfare-related events to settle arising disputes.[78]

This is not to suggest that the permissibility of arbitration in the commercial and financial context is contentious. On the contrary, the Council of *Fiqh* Academy–the Organisation of the Islamic Conference has generally accepted the permissibility of arbitration in commercial and financial disputes. It has also defined arbitration as 'an agreement between the disputing parties to appoint an arbitrator to settle the disputed matter with a binding decision that applies the principles of Islamic law'.[79]

This definition suggests that Islamic arbitration is not different from conventional arbitration as they both recognise the contractual basis of this dispute resolution mechanism. Therefore, the consent of the contracting parties, whether as a clause in the original contract or in a separate agreement, is the prime reason not only for the parties' engagement with the process but also for the binding character of the award.[80] Further, given the private nature of this mechanism, they both acknowledge that arbitrators are not judges. Yet they are always expected to be impartial, fair-minded, reasonable and knowledgeable with particular emphasis on their knowledge of Islamic law in the case of Islamic arbitration.[81] Finally, they both allow the parties to choose the law governing the disputes, which in the case of Islamic arbitration is Islamic law.

On the face of it, arbitration may seem an ideal mechanism to settle Islamic finance disputes; however, the reality is quite the opposite. It is argued that it would be a gross simplification to suggest that arbitration is a less complicated and more appropriate mechanism than litigation to resolve Islamic finance disputes. This argument is articulated in the context of England being the seat of arbitration.

In arbitration, the choice of law is not as straightforward as it may seem because there could be three types of law which are described as the applicable laws. First, is the curial law (*lex arbitri*) that applies to the arbitration procedures; second, the law governing the actual arbitration

77 Ibid verse [3:35] states: 'if you fear dissension between the two, send an arbitrator from his people and an arbitrator from her people'.

78 Wahba Al-Zuhayli, *Encyclopaedia of Islamic Jurisprudence and Contemporary Issues* vol 12 (Dar Al-Fikr 2012) 714.

79 Council of Fiqh Academy–Organisation of the Islamic Conference, Decision No 95/8/D9. Cited in Al-Zuhayli (n 78 above) vol 9, 611–612.

80 M L Moses, *The Principle and Practice of International Commercial Arbitration* 2nd edn (Cambridge University Press 2012) 2. Al- Zuhayli (n 78 above) vol 12, 731.

81 *Moses* (n 80 above) 2 and *Al- Zuhayli* (n 78 above) vol 12, 721.

clause or agreement in terms of its validity, scope and interpretation; and, third, the substantive law that governs the actual subject matter that is being arbitrated. It has been suggested that 'occasionally, but rarely' the curial law differs from the law governing the arbitration clause or agreement, while 'often' the substantive law differs from the first two.[82] This inherently creates a level of uncertainty and may open up the process in general, and the award for that matter, to a number of challenges. If the parties to an Islamic finance agreement choose England as the seat for arbitration, English law will apply and the main piece of legislation to consult in this respect is the Arbitration Act 1996.[83]

The curial law

Where parties have chosen England as the seat of arbitration, section 2(1) of the 1996 Act makes the provisions of part I of the 1996 Act applicable whether they explicitly chose English law as curial law or not. Part I includes the provisions that regulate the internal procedures of the arbitration (commencement of arbitral proceedings ss 12–14, the arbitral tribunal ss 15–29 and the arbitral proceedings ss 33–41, the award proceedings and rules ss 46–58) and the court's supervisory role at different stages of the arbitration (the court's power with regard to arbitral proceedings ss 42–45 and the court's power in relation to award ss 66–71).[84] Parties to an Islamic finance agreement may choose institutional arbitration where the conduct of the arbitration is supervised by a well-known international organisation that has its own set of procedural rules.[85] Yet, the parties still need to connect the conduct of the arbitral proceedings to a national legal system, such as English law (ie the 1996 Act). This is essential with regard to issues such as the extent of powers the parties have in selecting the arbitral procedures and the compulsory ones that the parties must adhere to, the national courts' involvement in the arbitration procedures and the review of awards.[86]

82 *European Film Bonds A/S v Lotus Holdings LLC* [2019] EWHC 2116 (Ch), para 136.

83 The focus of the analysis in this part of the article is England as the seat for arbitration. However, for some broad overview of arbitration practices in different countries, see Julio C Colon, 'Choice of law and Islamic finance' (2011) 46(2) Texas International Law Journal 411–436.

84 On the contrary, parties who choose to arbitrate in another country cannot, according to the 1996 Act, choose English law as the curial law, therefore, the provisions in part I of the 1996 Act will not apply: Arbitration Act 1996, s 2; see J Hill, 'Some private international law aspects of the Arbitration Act 1996' (1997) 46(2) International and Comparative Law Quarterly 295.

85 Such as the ICC Court of Arbitration or the London Court of International Arbitration.

86 Collins and Harris (n 38 above) vol 1, 16-009.

A close examination of what the 1996 Act is offering procedurally to the parties to an Islamic finance dispute demonstrates how disadvantaged this mechanism is compared to litigation.

First, the 1996 Act does not provide summary procedures where it imposes a duty on the arbitration tribunal to give each party 'a reasonable opportunity of putting his case and dealing with that of his opponent'.[87] While this clearly prevents the court's 'encroachment on the principle of party autonomy ... if the parties have agreed to arbitrate their dispute',[88] it reduces the efficiency of this mechanism and allows parties to play for time. For example, where the respondent does not have a real prospect of successfully defending their claim the claimant cannot have a summary decision in their favour.[89] Further, where one of the parties is a recalcitrant absentee from the hearing in arbitration, unlike litigation, a judgment in default of defence may not be entered into as of right without any examination of the merits.[90] Another example, where there is no genuine dispute to submit to arbitration, section 9 of the 1996 Act does not allow the parties to a disputed contract with an arbitration clause to obtain a summary judgment from the English court preventing the reference to arbitration.[91] It is worth noting that section 9 is mandatory: parties to an arbitration clause or agreement cannot contract themselves out of this provision.[92] Therefore, the lack of these summary procedures may unnecessarily prolong the process and allows either of the parties to use arbitration for tactical reasons.

Second, as for interim remedies or injunctions, in principle, these are designed to require one of the parties to act or refrain from acting in certain ways, which is central to the protection of the other party's rights. Given the sense of urgency that is associated with the use of these injunctions, their effectiveness depends on the following four factors: the breadth of these injunctions; their binding authority; the time needed to enforce; and the parties that they can be enforced against.

Section 38(1) of the 1996 Act deals with arbitral interim injunctions. It is not a mandatory section, therefore, parties can agree broad

87 Arbitration Act 1996, s 33(1)(a).

88 Lord Saville, 'The Denning Lecture 1995: arbitration and the court' (1995) 61(3) Arbitration 161.

89 D St J Sutton, J Gill and M Gearing, *Russell on Arbitration* 24th edn (Sweet & Maxwell 2015) 1-031.

90 R Merkin and L Flannery, *Merkin and Flannery on Arbitration Act 1996* 6th edn (Routledge 2020) 373. See also Arbitration Act 1996, s 33.

91 Arbitration Act 1996, s 9: see Lord Saville (n 88 above) 161 and N Blackaby, C Partasides with A Redfern and M Hunter, *Redfern and Hunter on International Arbitration* 6th edn (Oxford University Press 2015) 19.

92 Arbitration Act 1996, sch 1, Mandatory Provisions of pt I.

powers for the arbitral tribunal to grant interim injunctions or they can restrict these powers. However, if there is no agreement, then the 1996 Act provides default provisions, which are very limited, and the 1996 Act does not in any of its parts explicitly allow the court to interfere in the exercise of these powers.[93] These limited interim remedies include injunctions in relation to security of cost, preservation of property and preservation of evidence.[94] The limitation in the breadth of these default interim injunctions can be seen in relation to the preservation of property injunction. This injunction is of particular significance to Islamic finance agreements and any possible disputes arising. According to Islamic finance principles, money itself is not a commodity, and investing the capital in real property/assets is central to generating litigate profits, which makes the property/assets an integral component of any Islamic finance agreement.[95]

In this respect, it has been argued that the default preservation of property injunction is much narrower than that granted by the court as it only concerns a property subject to the proceeding and owned or possessed by a party to the proceeding.[96] This falls short of a court interim freezing injunction that covers assets/property not directly concerned with the proceedings and in control, rather than owned or possessed, by a party to the proceeding.[97]

Further limitations can be identified regarding the binding authority of the preservation of property injunctions, and the other default interim injunctions for that matter. They are granted in the form of orders or directions, therefore, the arbitral tribunal cannot compel compliance as it lacks the coercive power of the court, despite the urgency associated with their use.[98] Although they can be granted in the form of a peremptory order,[99] which can be enforced by the court under section 42 of the 1996 Act,[100] this requires time that defies the urgent nature of these interim injunctions. The party would first need to fail to comply with the interim injunction without showing sufficient cause, and only then could the tribunal issue the injunction

93 Merkin and Flannery (n 90 above) 398. See also Arbitration Act 1996, s 38.

94 Arbitration Act 1996, s 38(3), (4) and (6); see Sutton et al (n 89 above) 5-080, 5-081.

95 This can be seen in commercial agency agreement (*Wakalah*) and also mark-up agreement (*Murabaha*).

96 Arbitration Act 1996, s 38(4), and see also Merkin and Flannery (n 90 above) 404.

97 Pt 25 of the Civil Procedures Rules does not include these limitations. See further, Merkin and Flannery (n 90 above) 405.

98 Sutton et al (n 89 above) 5-082, 5-083.

99 Arbitration Act 1996, s 41.

100 Ibid s 42; Sutton et al (n 89 above) 5-083.

as a peremptory order.[101] This is not to suggest that the court does not have any powers in relation to interim injunctions in arbitration. Parties can seek urgent interim relief to preserve assets/property from the court under section 44(2)(e) and (3) of the 1996 Act.[102] However, on the one hand, the party seeking the relief will have to demonstrate its urgency in addition to the fact that the court's involvement would undermine the privacy and confidentiality advantage of arbitration.[103] On the other hand, similar to section 42 of the 1996 Act, section 44 is non-mandatory, which means parties to an arbitral clause or agreement can contract out of these sections.

It is worth noting that there is an argument for the use of section 39 of the 1996 Act to grant interim injunctions in the form of provisional awards. Section 39 empowers the tribunal to order on 'a provisional basis any relief which it would have power to grant in a final award', including, for instance, order for payment or the disposition of property between the parties.[104] However, there are a few issues concerning the use of section 39 in this context.

First, there is some doubt as to the interpretation of this section, given that only the title refers to the provisional measure as an 'award' while the section uses the term 'order'. This casts uncertainty in terms of the enforceability of the measure, given the difference in this respect between an 'order' and an 'award' albeit provisional.[105] Second, section 39 does not provide a more advanced interim injunction to preserve property than that available under section 38, as it does not allow the tribunal to grant interim freezing injunctions. Although there is no unanimity, a significant volume of the academic commentary on section 39 takes the view that a freezing order cannot be granted, as of right, in the form of a provisional award because it is not a relief the tribunal is empowered to grant in a final award under section 48 despite the broad wording of this section.[106] Rix LJ noted this view in the Court of Appeal decision of *Kastner v Jason*,[107] although it was

101 Arbitration Act 1996, s 42(3) and (4) and see also Sutton et al (n 89 above) 5-186.

102 Sutton et al (n 89 above) 5-083.

103 Merkin and Flannery (n 90 above) 470.

104 Arbitration Act 1996, s 39(2).

105 Especially in the context of domestic arbitration where orders can be granted in the form of a peremptory order (s 41) and enforced by the court (s 42) while an award albeit provisional can only be enforced under s 66. See Merkin and Flannery (n 90 above) 409.

106 Rix LJ in the Court of Appeal decision of *Kastner v Jason* [2004] EWCA Civ 1599, para (16), cited a number of academic sources, among which Sir M J Mustill and S C Boyd, *Mustill & Boyd, Commercial Arbitration* (LexisNexis 2001) 315, and Sutton et al (n 89 above) 6-020.

107 *Kastner* (n 106 above) paras 16–19.

'not the direct subject matter of any issue in this appeal'.[108] Third, and most importantly, this section needs to be agreed by the parties in the first place, as the 1996 Act does not set this power as a default option subject to contrary agreement.

As for the last indicator of the effectiveness of the arbitral interim measures, it is the parties that they can bind. In this regard, these interim injunctions can only bind parties to an arbitral clause or agreement, which excludes third parties. This is particularly disadvantageous in Islamic finance disputes as the majority of Islamic finance agreements involve a third party who would either sell or purchase the property/asset that is central to the finance agreement. Take, for example, a mark-up agreement (*Murabaha*) where the Islamic finance institution would buy a specific asset from a third party to resell it to the client at a mark-up price or appoint an agent to purchase the assets, as seen in the *Shamil Bank* case.[109] The third party in these examples could be in control of the asset/property and is not bound by the arbitral interim injunctions.[110]

The governing law of the arbitration agreement

It is more common for the agreement to arbitrate to be found as a clause in the main contract than as a submission agreement or a compromis.[111] In any case, identifying the law governing the arbitration agreement is a matter of significant importance as it decides on the 'validity, scope and interpretation of an arbitration agreement'.[112] Any successful challenge to any of these issues would adversely affect the validity and recognition of the final arbitral award and render it unenforceable.

Further, in the context of international commercial arbitration, identifying the law governing the arbitration agreement could also be a significant challenge where the parties have not made an express choice of law with specific reference to the arbitration agreement. The Rome I Regulation does not apply in this context;[113] therefore, where England is the arbitration seat, any questions arising before the English court concerning the governing law of the arbitration agreement will be

108 Ibid para (19). The first instance judge Mr Justice Lightman considered *obiter* that s 39 could be used to grant a freezing order where the parties conferred such powers in the tribunal on a final award, although in this case 'the arbitration agreement does not expressly grant to the Beth Din jurisdiction to grant a freezing direction in its final award': see *Kastner* (n 106 above), paras 27–28, and see also Merkin and Flannery (n 90 above) 410.

109 *Shamil* (n 11 above). The appeal decision is reported in [2004] 2 Lloyd's Rep 1.

110 In order for interim injunctions to bind third parties an application must be made to the court. Sutton et al (n 89 above) 5-080

111 Collins and Harris (n 38 above) vol 1 16-008.

112 Ibid 16R-001.

113 Art 1(2)(e) states that the Regulation does not apply to 'arbitration agreements'.

subject to English common law and the choice of law rules. This means that the court will decide whether the parties to arbitration made an implied choice, where there is no express choice, of the law governing the arbitration agreement and give effect to the parties' choice.[114] In the absence of such a choice, it is the law that is most closely connected to the arbitration agreement which would 'generally' be the law of the seat where the parties had chosen a seat of arbitration.[115] However, the application of these rules is not straightforward. The English court has taken different approaches. In some decisions the court has opted for the law governing the main contract chosen (expressly or impliedly) by the parties,[116] while in other decisions the court has opted for the law of the arbitration seat.[117] Further, views also differed as to whether either approach reflected the parties' implied choice or the closest connection test.[118]

This remained an area of uncertainty until the UK Supreme Court decision in *Enka Insaat ve Sanayi AS v ooo Insurance Company Chubb*.[119] The Supreme Court decided that in the absence of an express choice of law for the arbitration agreement the law chosen to govern the whole contract will apply to the arbitration agreement,[120] and where there is not such a choice then the law of the arbitration seat will apply to the arbitration agreement as the most closely connected to it.[121]

---

114 *European Film Bonds A/S v Lotus Holdings LLC* [2019] EWHC 2116 (Ch), para 142; see also Collins and Harris (n 38 above) vol 1 16R-001.

115 *Enka Insaat ve Sanayi AS v ooo Insurance Company Chubb* [2020] Bus LR 2242, 2293; and see also Collins and Harris (n 38 above) vol 1, 16R-001, fn 1 cites a large volume of case law decisions to this effect.

116 *Sulamérica Cia Nacional de Seguros SA v Enesa Engelharia SA* [2012] EWCA Civ 638; [2013] 1 WLR 102: see also *Black Clawson International Ltd v Papierwerke Waldhof-Aschaffenburg AG* [1981] 2 Lloyd's Rep 446, 456; *Channel Tunnel Group Ltd v Balfour Beatty Construction Ltd* [1993] AC 334, 357; *Arsanovia Ltd v Cruz City 1 Mauritius Holdings* [2012] EWHC 3702 (Comm) and *Habas Sinai Ve Tibbi Gazlar Istihsal Endustrisi AS v VSC Steel Co Ltd* [2013] EWHC 4071 (Comm) cited in fn 5 in Mark Campbell, 'How to determine the law governing an arbitration agreement: direction form the Supreme Court' (2021) 24(1) International Arbitration Law Review 29.

117 *Enak v Chubb* [2020] EWCA Civ 574. See also: *XL Insurance Ltd v Owens Corning* [2001] 1 All ER (Comm) 530; *C v D* [2007] EWCA Civ 1282 cited in fn 6 in Campbell (n 116 above). See also, William Day, 'Applicable law and arbitration agreements' (2021) 80(2) Cambridge Law Journal 239; see also Myron Phua and Matthew Chan, 'Persistent questions after Enka v Chubb' (2021) 137(Apr) Law Quarterly Review 217.

118 Day (n 117 above).

119 *Enka Insaat* (n 115 above).

120 Ibid 2292.

121 Ibid 2293.

Accordingly, in the context of Islamic finance disputes, where English law is identified as the governing law of the arbitration agreement, there are two issues that parties should take note of. First, the interpretation of the arbitration agreement and its scope are among the key areas that will be governed by common law, more specifically the general principles of contract law.[122] As seen in *Enka*,[123] in deciding that English law was the governing law of the arbitration agreement, that the arbitration clause was valid and that the claim fell within the scope of the arbitration agreement,[124] the Supreme Court engaged with some key common law contractual interpretation principles, such as the separability principle and validation principle.[125] Second, the Arbitration Act 1996 would also apply, as part of English law, to issues cornering formal validity of the arbitration agreement.[126] More importantly, the 1996 Act provides the arbitral tribunal with the freedom to determine its jurisdiction (s 30) and limits the court's power to determine the jurisdiction of an arbitral tribunal (s 32).[127] Yet section 67 still provides a route to challenge an award on jurisdiction in the English court.[128] While the 1996 Act does not dictate how the tribunal interprets these matters, it remains central to the parties to Islamic finance disputes to understand the powers given the tribunal in this respect, the challenges associated with them, and the routes to review these decisions by the English court.

The 1996 Act under section 30 empowers the arbitration tribunal, unless agreed otherwise, to rule on its own 'substantive jurisdiction'.[129] This includes the validity of the arbitration agreement, the constitution of the arbitration tribunal and the matters that fall within the arbitration agreement.[130] It has been suggested that this section is one of the most important sections of the 1996 Act as it deals with the difficult concept of 'substantive jurisdiction' and its categories.[131] In principle, this concept concerns the legal right or competence of the tribunal to

122 Campbell (n 116 above) 35.

123 *Enka* (n 115 above).

124 Ibid 2296.

125 Ibid 2256, 2261–2262, 2271–2276, see Campbell (n 116 above) 31–33. For more on the separability principle and construction of an arbitration clause, see also *Trust v Privalov* [2007] UKHL 40.

126 Arbitration Act 1996, s 5: see Collins and Harris (n 38 above) vol 1, 16-023–16-026.

127 Collins and Harris (n 38 above) vol 1, 16-013, and the source expands on this point in fn 29.

128 Ibid.

129 Arbitration Act 1996, s 30(1).

130 Ibid, s 30(1)(a), (b), (c).

131 Merkin and Flannery (n 90 above) 318.

decide over the matters put to it by the parties.[132] In practice, however, the concept of substantive jurisdiction remains elusive and has often been confused with the concepts of admissibility[133] and authority.[134] Even the categories identified by section 30(1)(a)–(c), especially the 'validity' of the agreement and the 'matters' submitted to arbitration, have been criticised for being either limited in coverage regarding the former or lacking clarity in relation to the latter.

The significance of the distinction between a jurisdictional and non-jurisdictional claim is that the former is subject to the English court review, and where the challenge is successful the award is rendered unenforceable. There are two sections under the 1996 Act, section 32 and section 67, which provide the route to the court review of jurisdictional challenges, and both sections are mandatory.[135]

As explained earlier, subject to parties' contrary agreement, the arbitral tribunal has the competence to decide on a question of its substantive jurisdiction. The decision of the tribunal under section 31(4), which is also a mandatory section, could be in a preliminary award, concerning only the jurisdictional issue, or in its final award, when deciding on the merits. The parties to arbitration are entitled to agree which course the tribunal should take in deciding on this issue.[136] Whether the jurisdictional question is decided by the tribunal in a preliminary or final award can be challenged before the English court under section 67.[137] Further, section 32 provides an alternative route to question the substantive jurisdiction of the arbitral tribunal before the court. This is not a form of appeal from an arbitral award on jurisdiction, but rather it is a substitute to that type of award.[138] The court can make a binding ruling on a jurisdictional matter if the application was made either with the agreement in writing of the parties,[139] or if it was permitted by the tribunal and the court is satisfied that this is likely to make substantial savings in costs, there

132 Ibid.

133 The best example is whether the time limit on claiming arbitration is a question that concerns the jurisdiction or the admissibility. If it concerns admissibility then it falls within the jurisdictional authority of the tribunal, therefore, it is not itself a jurisdictional issue that can be reviewed by the court and if successful annul the award. See Merkin and Flannery (n 90 above) 319–320.

134 Also acting within authority is not a jurisdictional matter that can be challenged before the court. However, in this case there is the argument of serious irregularity that can open it up to court review under s 68.

135 Arbitration Act 1996, sch 1, Mandatory Provisions of part I

136 Ibid s 31(4).

137 Ibid s 67(1)(a) in relation to an award confined to the jurisdictional matter or s 67(1)(b) in relation to a final award.

138 Merkin and Flannery (n 90 above) 359.

139 Arbitration Act 1996, s 32(2)(a).

was no delay in applying, and there was a good reason for referring the matter to the court.[140]

Given the earlier highlighted uncertainties associated with the concept of substantive jurisdiction and the unenforceability of an award in the case of a successful court challenge, it is argued that the parties to an arbitral Islamic finance dispute should always agree that the tribunal should decide the jurisdictional objection in an early ruling.[141] This is to save on time and costs that parties would incur if the jurisdictional objection was decided with the merits in the final award that later was successfully challenged before the court on a jurisdictional ground. For the same reason, parties also should always agree to stay the arbitral proceedings if an application to court was made under section 32. This is particularly important given that the default setting under the 1996 Act is that the arbitral tribunal 'may continue the arbitral proceedings' while an application under section 32 is pending.[142]

A case in point is *Al Midani & Another v Al Midani & Others*.[143] Although the case was decided according to the old arbitration laws, it remains a clear illustration of the significant impact that a successful court challenge to the jurisdiction of the arbitral tribunal will have on the enforceability of the award. It can be suggested that the case deals with a number of jurisdictional matters to which the court applied English law.

Article 7 of the arbitration agreement stated that any dispute arising from the arbitral decision concerning the distribution of the inheritance should be decided by 'an Islamic judicial body' appointed by the Trusteeship Council without the involvement of the heirs.[144] The Trusteeship Council invoked this article and appointed the Islamic Sharia Council in London to decide on the disputed arbitral decision. The claimants, two of the heirs, challenged the jurisdiction of the Islamic Sharia Council. The court in its decision dealt with a number of jurisdictional issues. First, was the constitution of the arbitral tribunal. The court found that the Islamic Sharia Council in London was neither a national court nor an arbitration tribunal, therefore, did not fall with the concept of 'an Islamic judicial body' as stated in article 7 of the arbitration agreement.[145] Second, was the validity of article 7 as an arbitration clause: the court found that article 7 was not an 'arbitral clause' or providing for a 'second tier of arbitration'.[146] Finally, even

140 Ibid s 32(2)(b)(i)–(iii).
141 Ibid s 31(4).
142 Ibid s 32(4).
143 [1999] CLC 904.
144 Ibid 906.
145 *Al Midani* (n 143 above) 906, 913.
146 Ibid 913.

if article 7 was an arbitration clause and the Islamic Sharia Council was an arbitral tribunal, the court found that it had no jurisdiction over the claimants. This is because only parties to the arbitration agreement (ie the heirs) can invoke the reference to 'an Islamic judicial body', which a representative of Trusteeship Council invoked in this case.[147] Accordingly, the decision of the Islamic Sharia Council was unenforceable.

The substantive law

In the years preceding the 1996 Act, the English court questioned, in the *dicta* of a couple of cases,[148] whether the rule that the substantive issues in an arbitration should be governed by the law of a country remained 'good law'.[149] The 1996 Act decisively settled this issue under section 46(1). It states that:

> the arbitral tribunal shall decide the dispute: (a) in accordance with the law chosen by the parties as applicable to the substance of the dispute, or (b) if the parties so agree, in accordance *with such other considerations* as are agreed by them or determined by the tribunal.[150]

The use of the term 'other considerations' by the 1996 Act in the context of the substantive law choice enables the parties to arbitration to choose the principles of a non-national system of law to govern the substance of their arbitrated dispute.[151] Therefore, parties to an Islamic finance arbitration can choose Islamic law to govern the substantive matters of their dispute.

Further, the 1996 Act allows mixing the principles of a non-national system of law with the principles of a national system of law.[152]

S*anghi Polyesters Ltd (India) v The International Investor KCFC*[153] put the application of section 46 to the test. Sanghi Polyesters Ltd (SPL), an Indian company, obtained finance (US$5 million) from the International Investor KCFC (TII) by using an Islamic finance agreement known as an '*Istisna*' or manufacturing agreement.[154]

147 Ibid 913–914.

148 *Lloyd LJ in Home and Overseas Insurance Co Ltd v Mentor Insurance Co* [1990] 1 WLR 153, 166; and *Deutsche Schachtbau Tiefbohr-Gesellschaft MBH v Shell International Petroleum Co Ltd* [1990] 1 AC 295, 315, cited in *Sayyed Mohammed Musawi v RE International (UK) Ltd* [2007] EWHC 2981 (Ch), para 21.

149 *Sayyed Mohammed Musawi* (n 148 above) para 21.

150 Arbitration Act 1996, s 46 (emphasis added).

151 Collins and Harris (n 38 above) vol 1, 16-053.

152 Hill (n 84 above) 300.

153 [2001] CLC 748.

154 Manufacturing agreement where the finance provider is contracted to manufacture the required product and then sells to the client (buyer) with an added margin of profit.

The contract of *Istisna* set a 9 per cent annual profit rate expected on this investment by TII. SPL defaulted on its agreed payments and TII claimed the repayment of US$5 million and the outstanding profits of US$230,417.

The parties agreed to an institutional arbitration (International Chamber of Commerce (ICC) arbitration) and chose London as the seat of arbitration. This meant that, in conjunction with the ICC rules, part I of the 1996 Act was applied as the curial law. The parties also provided that the substantive issue in the arbitration should be governed by 'the laws of England except to the extent it may conflict with Islamic Sharia, which shall prevail',[155] which makes Islamic law as the prevailing substantive governing law of the dispute. The agreement allowed TII to appoint a Sharia expert to advise the arbitrator on issues concerning Islamic law, whose advice would be binding on TII, SPL and the arbitrator. TII chose the appointed sole arbitrator Mr Samir Saleh, an experienced lawyer and a Sharia expert, to act in this capacity, a choice that was endorsed by SPL.[156]

The sole arbitrator decided that according to the substantive governing law, 'Islamic Sharia', TII was entitled to their principal and the outstanding agreed profits but he disallowed the payment of US$600,000 in damages as this would be Sharia non-compliant.[157]

SPL challenged the arbitral award before the English court 'on point of law' under section 69 and on the basis of 'serious irregularity' under section 68 of the 1996 Act.

First, as for the application of section 69, the court found that 'virtually all the issues of law complained of by SPL touch on the arbitrator's approach to Sharia law'.[158] Therefore, the court concluded that it has no jurisdiction to decide on matters related to 'Islamic Sharia', as according to section 82 and section 69 of the 1996 Act the court will only have jurisdiction with regard to questions of English

155 *Sanghi* (n 153 above) 750.

156 Ibid 750.

157 Ibid 749

158 The issues that SPL raised primarily concerned the legal nature and validity of the contract ibid 751–752.

law.[159] This demonstrates the English court's acceptance of Islamic law, a non-national system of law, as the substantive governing law upon which the arbitration award was based, without questioning its validity as a choice of law. The judge clearly stated that there was no 'need to characterise Sharia as a foreign law or code, or set of customs … whatever Sharia may be it is not the law of England and Wales'.[160]

It is worth noting that this also applies to other religious laws. In *Schwebel v Schwebel*,[161] the parties agreed to refer their dispute, concerning the distribution of inherence, to the Beth Din, the Court of the Chief Rabbi in London, where the substantive applicable law is Jewish law. The award was later appealed against before the English court in pursuance of section 69 of the 1996 Act. The court did not question the validity of Jewish law as a governing law and found that the arbitrators had applied Jewish law and that the court would only allow an appeal on a question of English law, which was not the case here.[162]

Second, with regard to the challenge of 'serious irregularity' under section 68 in *Sanghi Polyesters Ltd (India) v The International Investor KCFC*,[163] in principle, the English court repeatedly stressed that section 68 should not be used as a 'backdoor'[164] to question the

159 It is worth noting that the English court acceptance of a 'non-national system' to apply to an arbitral dispute was seen on occasions even before the 1996 Act in relation to other religious laws, such as Jewish law. The case of *Soleimany v Soleimany* [1999] QB 785 concerned an illegal contract to export Iranian carpets in contravention of Iranian revenue laws and export controls. A dispute arose in relation to the proceeds of a sale of illegally exported quantities of carpets from Iran. The parties signed an arbitration agreement to settle this dispute before the Beth Din in accordance with Jewish law. The case was brought before the English court with regard to the enforcement of this arbitral award in England. The key point to make is that the High Court did not question the validity of such an award, which was made on the basis of Jewish law, and granted leave to enforce the award. However, the Court of Appeal refused to enforce the award, not because of the invalidity of the choice of Jewish law, but because of the illegality of the disputed agreement in the first place, ie illegally exporting these carpets from Iran to be sold in the UK and elsewhere.

160 *Sanghi* (n 153 above) 751. It is worth noting that this also applies to cases where a foreign law is as the substantive governing law in arbitration. In *Egmatra AG v Marco Trading Corporation* [1998] CLC1552, 1552–1553, Swiss law was the governing law of the substance of the disputed matters. Therefore, the court dismissed the claim under s 69 as the question of law was not of the law of England and Wales.

161 [2010] EWHC 3280 (TCC).

162 Ibid para 14.

163 *Sanghi* (n 153 above).

164 *Warborough Investments Ltd v S Robinson & Sons (Holdings) Ltd* [2002] EWHC 2502 (Ch), para 4.

factual findings of the arbitral award,[165] whether the substantive applicable law is English, foreign laws or other considerations. Therefore, the test to apply this section sets the bar high where the irregularity must be 'serious' and 'has caused or will cause substantial injustice'. For example, the court in *JD Wetherspoon plc v Jay Mar Estates*[166] found that even if there was serious irregularity it had not in this case given rise to substantial injustice.[167] As in the case of *Sanghi Polyesters Ltd (India) v The International Investor KCFC*,[168] SPL's substantial ground of claim under section 68 was that the arbitrator in the award cited a few sources of Islamic law, which were never drawn to the attention of the parties before the award was delivered. The court found that there was no irregularity because, on the one hand, the fact that the parties agreed to give the sole arbitrator 'additional expert powers' to decide Sharia matters meant that the parties were 'clearly giving him more scope for individual initiative than is usual'.[169] More importantly, on the other hand, there was no injustice caused by the arbitrator to SPL.[170]

Based on the above, there are some key observations to make. First, the 1996 Act does not only validate the choice of a non-national system of law as the substantive governing law but it also protects the enforceability of the award based upon it. The court demonstrated that section 69 could only be used to question the award on a point of English law, which categorically excludes religious laws. Second, the court, time after time, rejected the use of section 68 as a backdoor to question the factual findings of the arbitrators. In other words, stopping any attempt to use 'serious irregularity' to involve the court in questioning how the arbitrator interpreted and applied the substantive law, which could be a non-national system of law, to the facts.

Having said that, there remains a major concern that stems from the nature of Islamic law, in its abstract concept, namely its certainty. Although the court's acceptance of Islamic law, as the substantive governing law, honours the autonomy of the arbitrating parties, which is the essence of arbitration, it does not add to the certainty of its application. As seen in *Sanghi Polyesters Ltd (India) v The International Investor KCFC*,[171] SPL questioned the validity of the

165 *Schwebel v Schwebel* (n 161 above) para 20, and see also *Lesotho Highlands Development Authority v Impregilo ApA* [2005] UKHL 43 (cited in ibid para 20).

166 [2007] EWHC 856 (TCC).

167 Ibid para 35. For more detailed commentary on s 68, see Merkin and Flannery (n 90 above) 693–730.

168 *Sanghi* (n 153 above).

169 Ibid 754.

170 Ibid.

171 Ibid.

underlying agreement as a genuine *Istisna*, and the parties' choice of 'Islamic Sharia' left it entirely to the arbitrator to decide what this meant. The court was in no position to assess the agreement's Sharia compliance nor was it able to displace Sharia and apply English law as it did in some of the litigated disputes discussed earlier. It can be suggested that the arbitrating parties would have improved the certainty of their choice if they added that the arbitrator's interpretation of Islamic Sharia, concerning the validity of the *Istisna* agreement, should be guided by AAOIFI Sharia Standard No 11,[172] as this standard provides parameters of a Sharia-compliant *Istisna* agreement. This is a clear point of reference that both parties know what to expect once applied.

Further, while the court expects the arbitrators' compliance with their duties under section 33 of the 1996 Act, there is no obligation on arbitrators to 'imitate the usual practice of an English judge'.[173] This has been demonstrated in the high threshold set by the court to successfully challenge an award on the basis of 'serious irregularity'. Therefore, parties who choose arbitration and Islamic law to govern the substance of the dispute should accept that arbitrators, while finding the facts, could not be held to the high standards expected of an English judge in litigation.

The final issue to consider in this context is the enforceability of an arbitral award based on Islamic law as the substantive law of the arbitration. In this regard, it has been suggested that 'arbitration is likely to afford the parties broad enforceability under the New York Convention on the Recognition and Enforcement of Foreign Arbitral Awards',[174] to which 169 countries have signed up.[175] However, there remains the question whether such an award could be challenged on the basis of public policy.[176] Given the focus of the article on English law, this issue will be considered in the context of the English court.

Article V(2)(b) of the New York Convention allows the court in the country where the recognition and enforcement is sought to set aside the award where it is contrary to the public policy in that country. The UK Arbitration Act 1996 (ss 100–103) re-enacts provisions from

172 AAOIFI, Sharia Standard No 11 (*Istisna* and parallel *Istisna*). See also White and King (n 61 above) 232.

173 Sanghi (n 153 above) 754.

174 White and King (n 61 above) 231.

175 New York Convention Contracting States as of 11 April 2022.

176 The interaction between Sharia and the concept of public policy has been examined in the literature in different contexts. Chuah, for example, analyses the enforcement of a foreign judgment where Sharia is a matter of public policy. See Chuah (n 27 above) 200–202.

previous legislation that implemented the New York Convention.[177] Accordingly, section 103(3) of the Arbitration Act 1996 makes public policy a ground to refuse recognition or enforcement of a New York Convention arbitral award.[178]

While all legal systems recognise the concept of public policy in the context of private international law,[179] the 'proper limits of the qualification of public policy remain elusive'.[180] However, Carter's analysis of when the English court invoked public policy in English private international law (conflict of laws system and enforcement of foreign judgments) drew some useful parameters for the concept of public policy. It has been suggested that public policy is fundamentally found on general principles of morality[181] where the court would refuse to apply a rule of a foreign law where it is 'unacceptably repugnant'.[182] Further, other grounds were found to be 'substantial justice' and the national and international interest of the UK.[183]

As for the meaning of public policy in international commercial arbitration, it has been suggested that what is relevant to the English court is the 'domestic' English concept of public policy.[184] Therefore, when it comes to the enforcement of a New York Convention arbitral award the concept of public policy encompasses 'the fundamental conceptions of morality and justice'.[185]

Accordingly, it is difficult to see how the enforcement of an arbitration award in an Islamic finance dispute that is based on Islamic law, as its substantive law, could be refused by the English court on a public policy ground. In *Sanghi Polyesters Ltd (India) v The International Investor KCFC*[186] the court enforced the arbitral award that according to Islamic law disallowed the payment of US$600,000 in damages to

177 Sutton et al (n 89 above) 8-025.

178 S 103(3) Arbitration Act 1996 states: 'Recognition or enforcement of the award may also be refused if the award is in respect of a matter which is not capable of settlement by arbitration, or if it would be contrary to public policy to recognise or enforce the award.'

179 Javier C De Enterria, 'The role of public policy in international commercial arbitration' (1990) 21(3) Law and Policy in International Business 390.

180 William E Holder, 'Public policy and national preferences: the exclusion of foreign law in English private international law' (1968) 17(4) International Comparative Law Quarterly 928–929.

181 P B Carter, 'The role of public policy in English private international law' (1993) 42 International and Comparative Law Quarterly 1–7.

182 Ibid 3.

183 Ibid 4.

184 Sir Jack Beatson, 'International arbitration, public policy, considerations, and conflicts of law: the perspectives of reviewing and enforcing courts' (2017) 33 Arbitration International 190.

185 Sutton et al (n 89 above) 8-050: fn 233 cites a large number of cases to this effect.

186 *Sanghi* (n 153 above).

prevent the circumvention of the prohibition of *riba*. Clearly the court did not find the award morally repugnant or in breach of substantial justice. On the other hand, *Soleimany v Soleimany*,[187] the dispute that concerned the proceeds of a sale of illegally imported quantities of carpets from Iran, was the subject of Beth Din arbitration, thereby, effectively Jewish law governed the substance of this dispute. The court refused to enforce the award, on a public policy ground, because of the illegality of the disputed agreement in the first place, namely illegally importing these carpets from Iran to be sold in the UK and elsewhere.[188] It is worth nothing that, although both cases did not concern New York Convention awards, they remain important in this context as both arbitral awards were based on religious laws as their substantive law, and the English court was asked to enforce the awards. Finally, the court repeatedly stressed that the use of public policy as a ground to refuse the enforcement of a New York Convention arbitral award should be used with 'extreme caution' as it was not intended to 'create an escape route'.[189]

CONCLUSION

When it comes to settling Islamic finance disputes, there is no binary choice to make between litigation and arbitration, where the latter is the optimal choice. On the contrary, arbitration is a complicated process that entails the application of multiple laws. As argued earlier, where the arbitration seat is England, the1996 Act does not provide the parties with the interim measures that are critical to protect the interests of the parties to a disputed Islamic finance agreement. Further, the question of the 'substantive jurisdiction' of the arbitrators remains a serious threat to the enforceability of the award. Finally, although the 1996 Act allows the choice of Islamic law, a non-national system of law, as the substantive governing law, this does not resolve a bigger problem concerning the certainty of its interpretation. The mere choice of an arbitrator who is expert in Islamic law does not necessarily guarantee a mutually accepted interpretation of Islamic law by all parties.

As for litigation, it is clear that English law is the most effective choice of governing law before the English court, which is not always equipped to serve the Sharia aspect of the disputed agreement. Nevertheless, this article has demonstrated that there is a way forward under the English private international law framework that allows the parties to benefit from the high standards of the English judiciary, its

187 *Soleimany v Soleimany* (n 159 above).

188 Ibid.

189 Sutton et al (n 89 above) 8-050: fn 232 cites a number of cases to this effect.

interim measures and enforceability of its decisions while remaining true to the Sharia essence of their agreement. The incorporation of the internationally accepted AAOIFI Sharia Standards in the terms of the agreement provides the court with a 'black letter' point of reference that the court can apply impartially.

Having said that, the extent to which such a solution can be utilised depends entirely on whether the parties are genuinely interested in the Sharia compliance of their agreement, and any arising disputes, or just the façade of it.

Northern Ireland
Legal Quarterly
Winter Vol. 73 No. 4 (2022) 650–684
Article DOI: 10.53386/nilq.v73i4.927

Creative Equity in practice: responding to extra-legal claims for the return of Nazi looted art from UK museums

Charlotte Woodhead*

Warwick Law School, University of Warwick
Correspondence email: c.c.woodhead@warwick.ac.uk

ABSTRACT

Looted cultural objects taken from Jewish owners during the Nazi Era still reside in museums worldwide. The United Kingdom's Spoliation Advisory Panel (the Panel) recommends solutions based on the moral strength of the claim where the original owner's legal title is extinguished. Using the framework of Equity this article argues that the Panel's work represents a modern, creative form of Equity. The Panel's work plugs a gap left by the law, much as Equity aimed to do. Despite a wide discretion to recommend just and fair solutions, the Panel is developing settled principles rather than applying inconsistent concepts of morality. This article's reconceptualisation of this process as firmly grounded in Equity enables the Panel's work to be more fully appreciated as *sui generis*. It may also enable the Panel to serve as a model for resolving other disputes about cultural objects.

Keywords: looted art; Nazi Era; Equity; cultural objects; personal property; limitation.

INTRODUCTION

Museums and private collections house cultural objects, collected across the generations and originally acquired in a whole host of different circumstances. These range from punitive missions where loot was captured, objects collected as part of scientific or archaeological missions and objects acquired during colonial times to objects that

* The original idea for this article stemmed from a discussion that I had years ago with Professor Dawn Watkins; when I described the work of the Panel, she commented that it was reminiscent of the development of the Court of Chancery. Since then this article has taken many forms and so thanks are due to Professors Andrew Johnston and Jonathan Garton for comments and advice on aspects of the current paper, as well as to Professors Rebecca Probert, Fiona Smith and Hugh Beale for detailed comments on previous iterations. An earlier version of this research was presented at *The New Work in Property Law* event at UCL and thanks are due for the helpful comments at that symposium, in particular from Professor Charles Mitchell and Dr Alison Dunn. The article has benefited from detailed comments by reviewers, to whom I am immensely grateful. As ever, all errors and omissions rest with the author.

were taken during the Nazi Era as part of Hitler's persecutory aims, many of which entered the art market and where museums and private collectors acquired objects, often unaware of their tainted provenance.

Claims relating to all of these cultural objects and calls to respond to past injustice are ever-present themes in the media. Vocal arguments are made for retaining museum collections intact, in particular framed within the so-called culture wars; for that reason, if a system is to be introduced which could facilitate rather than mandate the transfer of such cultural objects to communities of origin (in appropriate circumstances) it needs to be based on rational and consistent criteria providing proportionate and principled responses. The Spoliation Advisory Panel (the Panel), set up to hear claims relating to cultural objects taken during the Nazi Era, provides such a model.[1]

Whilst statutes of limitation perform necessary policy roles in civil law, frequently serving justice,[2] in the context of Nazi Era dispossessions of cultural objects these statutes extinguish legal claims for return in circumstances recognised both nationally and internationally as meritorious and worthy of resolution. Heirs therefore have what are widely considered to be just claims but have neither a legal claim nor a remedy. For that reason in 2000, responding to its international commitments, the United Kingdom (UK) Government established the Panel. This independent panel of experts is tasked with hearing claims against UK museums and galleries based on moral rather than legal grounds. The Panel's primary aim is to recommend just and fair solutions to the claims which can lead to museums returning cultural objects[3] to the heirs of the original owners who lost them during the Nazi Era. Thus, like Equity before it, the Panel plugs a gap left by the law.[4] Equity – capitalised – is used here to refer to 'the doctrines and remedies that developed from the equity jurisdiction of the Court of Chancery before it was abolished by the Judicature Acts of 1873

1 Thus lending further support to the view that the Panel's work can serve as a model for the development of similar processes to resolve claims for other cultural objects of which the original owners, communities or nations were dispossessed: Charlotte Woodhead, 'The changing tide of title to cultural heritage objects in UK museums' (2015) 22 International Journal of Cultural Property 229, 246–248. See generally Evelien Campfens, 'Restitution of looted art: what about access to justice' (2018) 2 Santander Art and Culture Law Review 185.

2 Discussed below at text to n 50.

3 The Panel has not defined the term 'cultural objects' but it has considered claims for objects that might be considered as traditional artworks such as paintings and also porcelain, collections of clocks and watches, manuscripts and an ivory Gothic relief. On no occasion has the Panel discussed whether an object was actually a cultural object.

4 As to this gap, see pp 661ff below.

and 1875' whilst 'equity' is treated as having a 'broader meaning corresponding to natural justice and morality'.[5]

Previous work has focused on how the Panel has, over time, developed principles on which to base its recommendations[6] and how legal principles can assist in making moral determinations[7] or the general parallels between alternative dispute resolution (ADR) and Equity.[8] The link between a broad notion of equity used by the French Holocaust restitution panel has also been made;[9] however, that article focused on the overall principle of equity (in part translating fair and equitable from French into English).[10] Recent arguments have been advanced preferring to frame the Panel's work within concepts of justice rather than morality, pointing to the 'inadequate', 'opaque' and 'ad hoc' nature of morality.[11] However, through close analysis of the specific recommendations of the Panel, this article develops the scholarship by clarifying the nature and scope of the moral strength to claims that emerge from these recommendations. The article thus frames the Panel's work as a form of Equity, evolving outside the courts,

---

5 J McGhee (ed), *Snell's Equity* 35th edn (Thomson Reuters 2020) 4. Following the Judicature Acts of 1873 and 1875, Equity and the common law were administered together in the courts.

6 Charlotte Woodhead, 'Nazi Era spoliation: establishing procedural and substantive principles' (2013) 18 Art Antiquity and Law 167.

7 Norman Palmer, 'Spoliation and Holocaust-related cultural objects: legal and ethical models for the resolution of claims' (2007) 12 Art Antiquity and Law 1 and Evelien Campfens, 'Sources of inspiration: old and new rules for looted art' in E Campfens (ed), *Fair and Just Solutions? Alternatives to Litigation in Nazi-Looted Art Disputes: Status Quo and New Developments* (Eleven International Publishing 2015) 37.

8 Thomas O Main, 'ADR: the new Equity' (2005) 74 University of Cincinnati Law Review 329. For some interesting analysis of the role that Equitable interests might play in restitution claims in the courts, see Elizabeth Pearson, 'Old wounds and new endeavors: the case for repatriating the Gweagal Shield from the British Museum' (2016) 21 Art Antiquity and Law 201 and Luke Harris, 'The Role of trusts in cultural property claims' (2017) 22 Art Antiquity and Law 1.

9 Claire Estryn, Eric Freedman and Richard Weisberg, 'The administration of equity in the French Holocaust-era claims process' in Daniela Carpi (ed), *The Concept of Equity: An Interdisciplinary Assessment* (Universitätsverlag Winter Heidelberg 2007).

10 Specifically, it was critical of the way in which decisions had been reached for Holocaust claims but did not analyse the work of the CIVS (Commission pour l'indemnisation des victimes de spoliations) from the specific point of its wider role as an equitable forum and did not, understandably, situate it within the context of English Equity.

11 Debbie De Girolamo, 'The conflation of morality and "the fair and just solution" in the determination of restitution claims involving Nazi-looted art: an unsatisfactory premise in need of change' (2019) 26 International Journal of Cultural Property 357, 362.

but having at its heart the aim of achieving just and fair solutions by assessing the moral strength of claims.

Close parallels between the Panel and the justification for, development of, and remedies found in the early development of Equity exist.[12] Making these similarities visible and analysing their role in this article serves to legitimise this process as a *quasi*-legal one making use of Equitable principles. Rather than responding according to the caprice of its members, the Panel responds to claims in a measured and considered manner, developing principles – some of which share similarities with Equity, others which are *sui generis*, but all of which can serve as a model for future claims for cultural objects claimed by other communities and, in turn, legitimise those processes.[13] The strong public support for the Panel's work and its internal accountability based on situating it within a framework of a Creative Equity provides an external legitimisation which assists in transposing this process to other types of claim.[14]

The argument is not being made that the Panel's work is the same as Equity, but, instead, the central thesis of this article is that there are clear parallels and analogies that can be drawn with the role that Equity played in the development of Equitable principles to respond to a gap in the law and to deal with potential injustices. The Panel has developed a similar body of principles to Equity which it applies to these important intergenerational claims. The importance of drawing these analogies is to provide the normative justification for the existence of the Panel's process itself but also in terms of presenting a model to inform the process of return of other contentious cultural objects, such as those acquired during colonial times, as well as a result of punitive expeditions.[15] To this end, an interpretative analysis of aspects of the Panel's work through the lens of Equity is taken, drawing parallels

12 The primary focus is on the early development of the jurisdiction of Equity, from the petitions to the Chancellors to the Court of Chancery before the Judicature Acts.

13 Here the term 'communities' is used to refer to indigenous communities, cultural communities and nations which seek to claim cultural objects currently residing in museums.

14 With thanks to Professor Andrew Johnston for helping make visible this argument. Regarding matters of procedural legitimacy of the Panel and its European counterparts through the framework of Luhmann, see Matthias Weller, 'Key elements of just and fair solutions – the case for a restatement of restitution principles' in Campfens (n 7 above) 207.

15 As to which see Dan Hicks, *The Brutish Museums* (Pluto Press 2020).

between the two and constructing an image of Equity in a modern, *quasi*-legal setting.[16]

The first part of this article sets out the background leading to the Panel's establishment as a forum in which to hear claims based on moral as well as legal arguments. The second part analyses the Panel's role as a form of individualised justice which plugs a gap left by the law, drawing parallels with the reason why Equity found its place within the legal system. The third part explores the nature and scope of the claims heard by the Panel while the fourth focuses on the creativity of the Panel's remedies and its seemingly wide discretion, drawing parallels with Equitable remedies and the use of discretion when responding to claims.

THE PROBLEM: NAZI ERA DISPOSSESSIONS

Art and cultural objects were displaced during the Nazi Era on an unimaginable scale.[17] The circumstances in which the dispossessions took place were various, including direct seizure,[18] forced sales to fund exit visas,[19] or to pay tax bills demanded purely because their owners

16 The notion of *quasi*-legal is explored below at text to n 95. It is acknowledged that, as Lord Simon of Glaisdale observed in *United Scientific Holdings Ltd v Burnley Borough Council* [1978] AC 904, 944, modern-day Equity is administered in a fused system. See also J H Baker, *An Introduction to English Legal History* 4th edn (Butterworths 2002) 114; it therefore has a secondary or supplementary role to law (*Dudley v Dudley* (1705) Prec Ch 241, 119; F W Maitland, *Equity: A Course of Lectures* (Cambridge University Press 1936) Lecture II, 19). Nevertheless, in the context of the Panel's work, Equity is clearly at the heart of both its process and substantive recommendations and thus has an enhanced, rather than supplementary role as found in the fused system of the modern-day courts. To adopt the metaphor used by Worthington to describe the integration of Equity into the law (S Worthington, *Equity* 2nd edn (Oxford University Press Clarendon Law Series 2006) 32), here Equity has been integrated into the fabric of the Panel which in turn derives much from the law. For that reason, the focus in this article is firmly placed on the usefulness of drawing on concepts from Equity.

17 Seventh Report of the Select Committee on Culture, *Media and Sport, 'Cultural Property: Return and Illicit Trade* (HC 1999–2000 371-I 000) para 169.

18 Eg Report of the Spoliation Advisory Panel in respect of three pieces of porcelain now in the possession of the British Museum, London, and the Fitzwilliam Museum, Cambridge (11 June 2008) (2008 HC 602) (*Rothberger* claim), Report of the Spoliation Advisory Panel in respect of three drawings now in the possession of the Courtauld Institute of Art (24 January 2007) (2007 HC 200) (*Courtauld/Feldmann* claim) and Report of the Spoliation Advisory Panel in respect of four drawings now in the possession of the British Museum (27 April 2006) (2006 HC 1052) (*British Museum/Feldmann* claim).

19 Recommendation regarding Berolzheimer, RC 1.166, Dutch Restitutiecommissie, 4 September 2017.

were Jewish.[20] Other transactions appeared legal on their face but concealed sales necessitated by the circumstances of persecution.[21] In some situations persecuted owners sold valuable art or cultural objects whilst fleeing the Nazis,[22] or once they had reached safety in non-occupied countries (but would have been unlikely to have sold them but for the persecution).[23] During the war the Allies confirmed their commitment to untangling these varied transactions.[24] After the war significant efforts were made to return the displaced cultural objects;[25] in many cases the objects were returned to the countries from which they originated rather than directly to the individual owners – often because they could not be identified.[26] Some claimants received post-war compensation through the German claims process.[27]

The scale of the problem of dispossessed cultural objects was less in the UK, not having been an occupied country, and so the state was not in receipt of large collections of art and cultural objects of unknown ownership at the end of the war. Nevertheless, both national and non-national museums, as well as private collectors, purchased objects

20 Report of the Spoliation Advisory Panel in respect of a painting now in the possession of Glasgow City Council (24 November 2004) (2004 HC 10) (*Glasgow City/attrib Chardin* claim) and Report of the Spoliation Advisory Panel in respect of a painted wooden tablet, The Biccherna Panel, now in the possession of the British Library (12 June 2014) (2014 HC 209) *(British Library/Biccherna* claim).

21 These sorts of transactions were envisaged as early as 1943 in the Inter-Allied Declaration against Acts of Dispossession Committed in Territories Under Enemy Occupation or Control, London, 5 January 1943.

22 Eg Report of the Spoliation Advisory Panel in respect of a painting now in the possession of the Tate Gallery (18 January 2001) (2005 HC 111) (*Tate/Griffier* claim).

23 Report of the Spoliation Advisory Panel in respect of fourteen clocks and watches now in the possession of the British Museum (7 March 2012) (2012 HC 1839) (*British Museum/Koch* claim). These are known as *Flughtgut*.

24 Inter-Allied Declaration (n 21 above). Although this document informed the time-limited military laws in the different sectors, it was not incorporated into English law.

25 Robert M Edsel with Brett Witter, *The Monuments Men: Allied Heroes, Nazi Thieves, and the Greatest Treasure Hunt in History* (Preface 2009).

26 These became the MNR collection in France: Musées Nationaux Récupération: and the NK in the Netherlands: Nederlands Kunstbezit collection.

27 See, for example, the cases of Report of the Spoliation Advisory Panel in respect of eight drawings now in the possession of the Samuel Courtauld Trust (24 June 2009) (2009 HC 757) (*Courtauld/Glaser* claim) and Report of the Spoliation Advisory Panel in respect of an oil painting by John Constable, 'Beaching A Boat, Brighton', now in the possession of the Tate Gallery (26 March 2014) (2014 HC 1016) (*Tate/Constable* claim) [54].

after the war which had a tainted Nazi Era provenance.[28] Standards of provenance research undertaken when acquiring works in the 1950s and 1960s (when many objects entered museum collections) were less stringent than nowadays.[29] Whilst there is now a widespread appreciation of the potential for an object to have a tainted provenance where there are gaps in information between 1933 and 1945, previously this would not in itself have been a red flag dissuading the museum from acquiring the object.[30]

For several reasons the late 1990s saw renewed efforts to resolve claims against current possessors.[31] First, Holocaust restitution had become a more widespread topic, with class action claims and the establishment of the Jewish Claims Conference to deal with claims for lost life insurance policies, slave labour claims and for gold, primarily held by Swiss banks.[32] Claims for art did not lend themselves to class actions given the individual nature of works[33] and the ability to identify them in museum collections or at auctions. Yet, the claims for cultural objects were situated within this renewed interest in restitution, and were aided by the opening of previously sealed Eastern European archives[34] and the academic research which set out the scale of the

28 For the scale of the objects in museums with unknown provenance between the years 1933–1945, see the Collections Trust: Spoliation Research by UK Museums.

29 *Rothberger* claim (n 18 above) [14].

30 In its recommendations the Panel has not criticised omissions by museums relating to provenance during the 1950s and 1960s. However, the Panel adopted a different approach where an object was acquired in the 1980s (*Tate/Constable* claim (n 27 above)), although in claims after 2016 the actions of the museum when acquiring the object are unlikely to be subjected to such scrutiny since any moral obligation on the museum is now only considered if the Panel 'finds it necessary to do so to enable it to arrive at a fair and just recommendation' (Spoliation Advisory Panel Constitution and Terms of Reference (SAP ToR) [16]) rather than as one of the factors that the Panel should consider when determining a claim (as was the case under para 7(g) on the original Terms of Reference: *Spoliation Advisory Panel Constitution and Terms of Reference*, Hansard vol 348 col 256W (13 April 2000) [7(g)]).

31 Select Committee Seventh Report (n 17 above) [179]–[183].

32 Conference on Jewish Material Claims Against Germany.

33 See M Dugot, 'The Holocaust Claims Processing Office: New York State's approach to resolving Holocaust era art claims' in M J Bazyler and R P Alford (eds), *Holocaust Restitution: Perspectives on the Litigation and its Legacy* (New York University Press 2006) 279.

34 Select Committee Seventh Report (n 17 above) [179]. See generally 'The barbarians of culture' in S E Eizenstat, *Imperfect Justice: Looted Assets, Slave Labor and the Unfinished Business of World War II* (Public Affairs 2003) ch 9.

issue.[35] Additionally, there were strong international commitments to dealing with these claims including the Principles with Respect to Nazi-Confiscated Art concluded at the Washington Conference on Holocaust-Era Assets in 1998[36] and the Council of Europe's Resolution 1205.[37] Clear political support for returning these objects to their original owners was also made at the national level.[38] At this time, many museums in the UK and abroad engaged in detailed collections research and identified objects with gaps in their ownership history between 1933 and 1945.[39]

Despite this renewed interest, there was no flurry of litigation in the UK or across Europe because legal claims were unlikely to succeed. Although there are statements of English law indicating that the courts will refuse to uphold a transfer effected through foreign confiscatory legislation,[40] this has not been tested in the context of claims for Nazi Era dispossessed cultural objects. In *Oppenheimer v Cattermole* Lord Cross said 'what we are concerned with here is legislation which takes away without compensation from a section of the citizen body singled out on racial grounds all their property on which the state

35 Eg Lynn Nicholas, *The Rape of Europa* (Macmillan 1994); Hector Feliciano, *The Lost Museum: The Nazi Conspiracy to Steal the World's Greatest Works of Art*, T Bent (trans) (Basic Books 1997); and Jonathan Petropoulos, *The Faustian Bargain: The Art World in Nazi Germany* (Allen Lane 2000). See also Eizenstat (n 34 above).

36 Washington Conference Principles on Nazi-Confiscated Art.

37 This is particularly strong where the objects are in publicly funded institutions, where return is prioritised to avoid perceptions that the public are benefiting from tainted cultural objects: Council of Europe Resolution 1205 On Looted Jewish Cultural Property (November 1999), principle 12.

38 Eg the creation of the Mattéloi Mission working party on spoliation in France which was established by the French Prime Minister in 1997, the Origins Unknown Committee (the Ekkart Committee) in the Netherlands which was established by the Dutch Government in 1998, the Spoliation Working Group established by the UK's National Museum Directors' Conference in 1998 and in Austria the enactment of the 1998 restitution law, BGB1 Nr 181/1998, amended in 2009.

39 See n 28 above and Statement of Principles and Proposed Actions on Spoliation of Works during the Holocaust and World War II Period (National Museum Directors' Conference, London 1998). See Jacques Schumacher, 'British museums and Holocaust-era provenance research' in Ruth Redmond-Cooper (ed), *Museums and the Holocaust* (IAL Publishing 2021).

40 *Oppenheimer v Cattermole* [1976] AC 249 (HL) 276 and 268 (Lord Cross) *obiter*. This is set within the context of the Radbruch formula: Gustav Radbruch, 'Statutory lawlessness and supra-statutory law (1946)' (2006) 26 Oxford Journal of Legal Studies 1, 7.

passing the legislation can lay its hands on'.[41] That case involved taxation and the removal of German citizenship from Jewish people during the Nazi Era. This principle has not been applied to private transactions involving individual property rights or where there have been subsequent property transfers. Certainly, where an innocent third party has purchased the object, courts would be reluctant to unravel later transactions particularly where the current possessor had legal title and the original owner's claim and title were extinguished by the Limitation Act 1939.[42] Under this Act, time starts to run even in favour of a thief, whereas the reforms brought about by the Limitation Act 1980 mean that for theftuous conversions time only starts running from the first good faith purchase unrelated to the theft,[43] and never in favour of a thief or a convertor related to the theft.[44] A further shortcoming of both common law and Equity (until 2010)[45] was that the governing statutes of many national museums[46] restricted transfers from their collections,[47] even if justified on moral grounds.[48] However, Parliament intervened and the trustees of the national museums now have the power to transfer objects of which the owners lost possession during the Nazi Era where the Panel has recommended return and the Secretary of State has approved it.[49]

---

41 *Oppenheimer* (n 40 above) 268 (Lord Cross) *obiter*. See also F A Mann 'The present validity of Nazi nationality laws' (1973) 89 Law Quarterly Review 194, 205, referring to the existence of situations where it would be 'contrary to the judicial conscience, authority and dignity to give effect to an enactment which shocks one's sense of propriety and morality'.

42 Limitation Act 1939, s 3(2), or Limitation Act 1980, s 3(2). See Ruth Redmond-Cooper and Charlotte Dunn, 'Original but not enduring title: issues of space and time' in Ruth Redmond-Cooper (ed), *Museums and the Holocaust* 2nd edn (IAL Publishing 2021) 18.

43 Limitation Act 1980, s 4(2).

44 Ibid s 4(3).

45 When the Holocaust (Return of Cultural Objects) Act 2009 came into force; this provides an exception in the case of Nazi Era dispossessions to the prohibitive statutory provisions.

46 That is museums governed by statute in receipt of direct government funding: *Loans between National and Non-National Museums: New Standards and Practical Guidelines* (National Museum Directors' Conference, London 2003).

47 Eg British Museum Act 1963, s 3(4)), and Museums and Galleries Act 1992, ss 4(3), (4), (5) and (6).

48 These restrictive governing statutes even prevented the use of the principle in *Re Snowden* [1970] Ch 700 (Ch) which allows the payment of an *ex gratia* sum from charity property with the permission of the Attorney General: *AG v Trustees of the British Museum* [2005] Ch 397 (Ch) [45] (now see Charities Act 2011, s 106).

49 Holocaust (Return of Cultural Objects) Act 2009, s 2. This power is now an indefinite one: s 4 (as amended by the Holocaust (Return of Cultural Objects) (Amendment) Act 2019, s 1).

Statutes of limitation have important policy justifications,[50] including the desire to avoid potential injustice to defendants[51] and the risks involved with relying on stale evidence.[52] In the context of the former justification, it has been said that 'Long dormant claims have often more of cruelty than of justice in them.'[53] Nevertheless, leaving unresolved claims that occurred in extreme circumstances, often coupled with genocide and where the objects are of importance to modern-day claimants, may have more of cruelty than of justice in them and justify circumvention of these rules; certainly this is supported by international commitments.[54] Given the potential legal difficulties of retroactively imposing laws which would have had the practical effect of reviving otherwise extinguished property rights and, in turn, interfering with the property rights of the current possessors, more creative solutions were required.

JURISDICTION AND PROCEDURE: A NEW FORUM AND A NEW CLAIM

The UK Government was clearly committed to responding to this problem, but did not introduce legislative changes to extend or disapply the limitation periods; instead, it appointed an independent panel of experts to hear claims from those who lost possession of cultural objects during the Nazi Era (or from their heirs) which are now in a national museum or other museum or gallery established for the public benefit.[55] The Panel can also advise parties where an object is owned by a private collector at the joint request of both parties.[56] Claims from people whose legal claims would otherwise be time-barred therefore have a forum within which to be heard. Although the Panel does not determine legal title it investigates the original title of the owner and the museum's current title.[57] The Panel's 'paramount purpose' is to 'achieve a solution which is fair and just both to the claimant and the

50 *Board of Trade v Cayzer, Irvine and Company Ltd* [1927] AC 610, 628 (Lord Atkinson) and the Law Commission, *Limitation of Actions: Consultation Paper* No 151 (2013) paras 1.22–1.38.

51 *Abdulla and Others v Birmingham City Council* [2013] 1 All ER 649 (SC), 666 (Lord Sumption).

52 Ibid.

53 *A' Court v Cross* (1825) 3 Bing 329, 332–333 (Best CJ).

54 Eg Council of Europe Resolution (n 37 above) principle 13.1.

55 *Spoliation Advisory Panel Constitution and Terms of Reference*, Hansard (n 30 above); revised SAP ToR (n 30 above) [1].

56 SAP ToR (n 30 above)[6]. To date no such claims have been considered by the Panel.

57 Ibid [8].

institution'.[58] This underlying principle, derived from the Washington Conference Principles,[59] is mirrored in the approaches of similar panels established across Europe.[60] The approach taken here is to focus on the work of the UK's Panel through the lens of Equity in order to demonstrate the internal consistency of the process and with a view to framing specifically the nature of the claim and the remedies.[61] The Panel responds to the moral imperative to right historical wrongs and to return cultural objects to their 'rightful owners'.[62]

Although the Panel's recommendations have no legal force[63] and its proceedings are not a process of litigation,[64] claimants who accept the Panel's recommendations are expected to do so in full and final settlement of the claim.[65] The Panel's recommendations have been followed by the parties, save for a few isolated situations where particular legal impediments to the proposed solutions existed.[66] In part, museums may follow the recommendations because of the risk otherwise of professional embarrassment.[67]

Equity may be described as 'an instrumentality by which the adaptation of law to social wants is carried on'[68] and can mean 'any body of rules existing by the side of the original civil law, founded on distinct principles and claiming incidentally to supersede the civil

58 Ibid [14].

59 Above n 36.

60 Eg Die Beratende Kommission im Zusammenhang mit der Rückgabe NS-verfolgungsbedingt entzogener Kulturgüter, insbesondere aus jüdischem Besitz (Germany); Commission pour l'indemnisation des victims de spoliations (France); the Adviescommissie Restitutieverzoeken Cultuurgoederen en Tweede Wereldoorlog (Netherlands); and Der Kunstrückgabebeirat (Austria).

61 There are similarities in the approach of the committees across Europe and at times inconsistencies of outcome of claims involving the same claimants outside the scope of this article, which seek to use the UK Panel's process as a case study for conceptualising of the role, procedure, nature of the claim and the remedies as having close parallels with Equity. These differences are being explored by the 'Restatement of Restitution Rules for Nazi-Confiscated Art' Research Project.

62 Select Committee Seventh Report (n 17 above) [193].

63 SAP ToR (n 30 above) [10].

64 Ibid [9].

65 Ibid [11]. Palmer suggested that where the parties have accepted the recommendations of the Panel then this should act as an estoppel: Norman Palmer, 'The best we can do? – Exploring a collegiate approach to Holocaust-related claims' in Campfens (n 7 above) 179.

66 See Charlotte Woodhead, 'Putting into place solutions for Nazi Era dispossessions of cultural objects: the UK experience' (2016) 23 International Journal of Cultural Property 385, 396–397.

67 Ibid 395.

68 Henry Sumner Maine, *Ancient Law: Its Connection with the Early History of Society, and its Relation to Modern Ideas* (Cambridge University Press [1816] 2012) 28.

law in virtue of a superior sanctity inherent in those principles'.[69] This wider approach reflects the moral origins of equity but also the supplementary nature of Equity's relationship with the common law.[70] The Panel's work is clearly a mechanism to achieve the social wants of achieving just and fair solutions for Nazi Era victims. Where the parties follow the recommendations, the civil law rights of the respondent are superseded; if a national museum upholds the Panel's recommendation and transfers an object this has the significant effect of acting as an exception to the statutory prohibition on transfers from their collections.[71] The Panel's underlying ethos is to provide a means of resolving disputes for losses during genocide, recognised as both a national and international imperative.[72] Achieving justice (through just and fair solutions) is the superior sanctity inherent in the principles applied by the Panel.

Plugging a gap left by the law[73]

The Panel provides the only realistic forum in which to hear claims for Nazi Era dispossessions given the restrictive effect of limitation periods.[74] In all claims heard by the Panel to date the respondent has always had the best legal title to the object since the original owners' title and claim have been extinguished by the Limitation Act 1939. Therefore, all claims have been firmly based on moral considerations. As a separate forum from law, the Panel mirrors the work of the Chancellor and later the Court of Chancery where petitions were made by plaintiffs unable to fit claims within the prescribed forms of common

69 Ibid.

70 As to which see above n 16.

71 See above n 45.

72 See above at text to n 36 and n 37.

73 Watt describes the equity gap as the gap 'between the general law and more pleasing justice': Gary Watt, *Equity Stirring* (Hart 2009) 10. Miller also observes that Equity 'supplements the law by filling gaps that, for one or reason or another (and perhaps purely by chance), have not been filled otherwise': Paul B Miller, 'Equity as supplemental law' in Dennis Klimchuk, Irit Samet and Henry E Smith (eds), *Philosophical Foundations of the Law of Equity* (Oxford University Press 2020) 102.

74 See discussion above at text to n 42. A claimant would only have an arguable case if fraud or concealment were established which might disapply the effect of the limitation period: Limitation Act 1939, s 26.

law writ,[75] such that Chancery was thus 'to soften and mollify the Extremity of the Law'.[76] Whilst the early Court of Chancery mitigated the harshness of the common law, the Panel mitigates the harshness of both the common law and Equity. For without the Panel's intervention even the discretion permitted by Equity to forego charity property would not facilitate a transfer from a national collection.[77] Not only would legal claims have little chance of success, but resorting to ADR would not be possible in a claim against a national collection[78] as the power to transfer only arises where the Panel recommends return and the Secretary of State approves this.[79] Therefore, even if the parties had agreed to a mediated settlement to transfer an object, transfer would not be permitted unless the Panel first recommended it.[80] The Panel therefore provides a *substitute* process in a new forum (with a wholly different form of claim), rather than an *alternative* process to formal adjudication to resolve an extant legal claim. In practice, it overcomes the problem of the claimant's lack of legal title to the object but also, since 2010, plays a vital role in circumventing the problem of restrictive governing statutes of the national museums. The Panel's work therefore reflects the way in which Equity was conceived of as 'a moral virtue' which 'qualifies, moderates and reforms the rigour,

75 Given that the actions of men 'are so divers and infinite, That it is impossible to make any general Law which may aptly meet with every particular Act, and not fail in some Circumstances' (*Earl of Oxford's Case in Chancery* (1615) Rep Ch 1 (1615) 21 ER 485) – the case described by Ibbetson as 'not really a report of a decision at all, but rather a justificatory essay on the nature of the Chancery's jurisdiction by the Lord Chancellor' (D Ibbetson, 'A house built on sand: equity in early modern English law' in E Koops and W J Zwalve (eds), *Law and Equity: Approaches in Roman Law and Common Law* (Brill 2013) 56). Baker (n 16 above) 102.

76 *Earl of Oxford's Case in Chancery* (n 75 above).

77 *AG v Trustees of the British Museum* (n 48 above).

78 Their governing statutes curtail transfers from their collections. Eg British Museum Act 1963, the Museums and Galleries Act 1992 and the National Heritage Act 1983. It is likely that forms of ADR would be suitable where claims are made against private individuals. The work of the Art Loss Register and Art Recovery International in recovering art for the victims of Holocaust dispossessions demonstrate that compromise of claims can be achieved and presumably methods of ADR could be used. Both Christie's and Sotheby's auction houses have restitution policies and seek to resolve claims: see Christie's Restitution Services and Sotheby's Art Restitution.

79 Holocaust (Return of Cultural Objects) Act 2009, ss 2(2) and (3). Governing bodies will not necessarily rubberstamp the recommendation, although they may be under professional and public pressure to do so.

80 It would only be in the case of a genuine compromise of a legal claim that the trustees might be able to transfer an object without recourse to the Panel: *AG v Trustees of the British Museum* (n 48 above) [28]. This would depend on the claimant having an enduring legal title to the object: see generally Pearson (n 8 above) 212.

hardness, and edge of the law'[81] where 'such as have undoubted right are made remediless'.[82]

Yet, the establishment of the Panel demonstrates something more than simply an alternative place to hear an existing category of claim – instead, it is a new place to hear a different type of claim.[83] The claims heard by the Panel include claims that would likely have succeeded in law were it not for the fact that the Limitation Act had extinguished title and prevented the claim; such examples would be where there has been theft[84] and there would be a civil action in the tort of conversion.[85] These are therefore in some sense legal claims that are resurrected as moral claims and are seen in the context of seizure of cultural objects by the Gestapo – for example in the *Rothberger* and *Feldmann* claims.[86]

A second category of claims relates to sales that were forced by duress; here there is again a legal basis to the original loss, but the fact that the current possessors are innocent third parties would mean that a legal claim in duress against them would not succeed (irrespective of the extinction of a claim because of the passage of time).[87] Therefore, the Panel's jurisdiction works not to give a second life to an extinguished legal claim, but rather to circumvent a bar on recovery that would have occurred and to facilitate a direct claim against a third-party museum.

The final type of claim that the Panel considers is *Flughtgut* or flight goods, as seen in the *British Museum/Koch* claim.[88] These moral claims considered by the Panel do not have an equivalent legal claim because these involve legitimate transfers of property that occurred in third countries but which were necessitated by the financial difficulties of the original owners. The factor that makes them susceptible to challenge before the Panel is that the original owners were in financial difficulties and had to sell their cultural objects because of Nazi persecution and were unlikely to have sold those objects but for the persecution that they suffered. These represent an entirely new type of claim within this moral jurisdiction.

In considering these three types of claims, it is possible to observe parallels between those claims in Equity which were capable of being

81 *Dudley and Ward (Lord) v Dudley (Lady)* (1705) Prec Ch 241, 244.

82 Ibid.

83 Specifically, a moral rather than a legal one.

84 Under the laws in the country where the loss happened.

85 A successful claim in conversion could happen where the original act of conversion was abroad; the Panel has considered claims where losses took place in Austria or the then Czechoslovakia: *Rothberger,* and *Courtauld/Feldmann* and *British Museum/Feldmann* claims respectively (n 18 above).

86 Above n 18.

87 See *White v Garden* (1851) 10 CB 918 and Lord Cairns LC in *Cundy v Lindsay* (1878) 3 App Cas 459, 464.

88 Above n 23.

legal, but for certain formalities,[89] and the entirely new creatures of Equity such as restrictive covenants.[90] Thus, like the Court of Chancery which was a forum able to hear a new type of claim – an Equitable one – the Panel can hear a new claim based on the moral strength of the claim, seeking a just and fair solution. In both situations law fails the justice of the case, but in turn, Equity and the Panel provide new doctrines which permit justice to be done, for 'judges of equity have always been ready to address new problems, and to create new doctrines, where justice so requires'.[91] Similarly,

> [t]he Panel is not an attempt to resurrect the law and the full, unbending panoply of legal process; nor should it be seen as such or judged as such. Rather, it is a unique and imaginative response to uniquely dreadful events.'[92]

Like Equity, it thus provides a 'second doorway to justice'.[93]

Unfortunately, whilst a doorway exists for all claimants for objects held in national museums and other museums and galleries established for the public benefit, it does not always lead to justice. Thus, in some ways the plug used to fill the equity gap is not fully watertight, for, in the case of objects held in private collections, the claim is only heard if the possessor, as well as the original owner (or their heirs), agree.[94]

Closer to Equity than equity

The Panel's work can be situated within a framework of a *quasi*-legal process, with Equity at its heart. This approach is justified on several bases. Despite the lack of legal force, the Panel's recommendations are *quasi*-legal in nature because they have a significant effect on matters of title to cultural objects which are held for the public benefit.[95] In practice, the Panel's recommendations have been followed and can result in a transfer of what is, in practice, part of an otherwise *de facto*

89 For example, Equitable leases as seen in *Walsh v Lonsdale* (1882) 21 Ch D 9.

90 *Tulk v Moxhay* (1848) 2 Ph 774.

91 *Westdeutsche Landesbank Girozentrale v Islington London Borough Council* [1996] AC 669, 696 (Lord Goff).

92 Sir Paul Jenkins KCB QC, *Independent Review of the Spoliation Advisory Panel* (DCMS 2015) [6.6].

93 A phrase used by Lord Neuberger to describe Equity in 'Equity, ADR, arbitration and the law: different dimensions of justice' (Fourth Keating Lecture, Lincoln's Inn 19 May 2010) [31].

94 The notion of claimants having a *de facto* claim against a national collection or other museum established for the public benefit is discussed in the next part since the agreement of the respondent museum is unnecessary in such claims. In the case of private owners there is therefore a greater role played by the auction houses and the art recovery companies in reaching settlements between previous owners and the current owners. See above n 78.

95 See Woodhead (n 6 above) 189.

inalienable collection. Here an analogy is drawn with the terminology of *quasi*-legislation found in both secular[96] and canon law which has a supplementary role, filling 'gaps in formal law'.[97] Supplementing the law is the key purpose of the Panel as it provides the only realistic forum, with the only type of claim with a realistic chance of success. Equity's supplementary role has been identified as threefold – filling gaps left by the common law, modifying or adjusting the common law by providing more appropriate remedies and adding distinctive legal forms.[98] These three features of Equity are closely mirrored by the Panel in terms not only of plugging the gap left by the law, but also in the creation of a new type of claim based on the moral strength of the claim as well as the more nuanced remedial responses provided by the Panel.

Furthermore, the Panel is more closely aligned to Equity as administered by the courts, rather than any broad notions of morality covered by the term 'equity' (in lower case),[99] for over time the Panel has developed its own procedural and substantive principles which have avoided applying unwieldy moral concepts or subjective and inconsistent recommendations.[100] It has effectively created its own processes and procedures akin to developing a limited form of precedent and, although not legally binding,[101] in a similar way to Equity it has set its processes within the context of English legal principles and processes.[102] It therefore has a *quasi*-legal nature within which Equity plays a significant role. In adopting a limited form of precedent, not only in terms of procedural matters, but also substantive ones, the Panel's development has mirrored that of Equity which initially took account

96 R E Megarry, 'Administrative quasi-legislation' (1944) 60 Law Quarterly Review 125, 126.

97 Norman Doe 'Ecclesiastical *quasi*-legislation' in N Doe, M Hill and R Ombres (eds), *English Canon Law: Essays in Honour of Bishop Eric Kemp* (University of Wales Press 1998) 95.

98 Miller (n 73 above) 102.

99 See discussion at n 5 above.

100 See generally Woodhead (n 6 above). Oost recognises that the structure of a legalistic paradigm is needed in addition to a moral paradigm for Nazi Era restitution committees (focusing on the UK and Dutch committees) to provide 'a certain predictability of proceedings': Tabitha Oost, 'Restitution policies on Nazi-looted art in the Netherlands and the United Kingdom: a change from a legal to a moral paradigm?' (2018) 25 International Journal of Cultural Property 139, 173.

101 SAP ToR (n 30 above) [10].

102 Report of the Spoliation Advisory Panel in respect of a painting held by the Ashmolean Museum in Oxford (1 March 2006) (2006 HC 890) [25].

of the 'course of Chancery'[103] and later developed into precedent with the advent of law reporting.[104]

Conceptualising the Panel as a *quasi*-legal process dispensing Equitable principles is more appropriate than situating it within ADR (which is the framework usually used). The Panel does not neatly fit into any of the usual categories of ADR.[105] However, using the terminology of ADR presupposes that there is an extant legal claim for which the court would be a possible forum for the resolution of the dispute but where an alternative method (outside litigation) is sought. Yet in all claims so far the claimants have had no extant legal rights; the Panel thus provides an entirely different system in which to hear a very different type of claim.[106]

A further point of divergence from ADR is that rather than being a consensual means of resolving a dispute, the Panel has a *de facto* jurisdiction in certain situations. Specifically, where a claim is made for an object in a national museum or other museum established for the public benefit, a panel is convened to hear the claim and the respondent

---

103 *Barkley (or Berkley) v Markwick and Others* (1617) Ritchie 14, 15.

104 In *Gee v Pritchard* (1818) 2 Swanston 403; 36 ER 670, 674 Lord Eldon LC said: 'The doctrines of this Court ought to be as well settled and made as uniform almost as those of the common law, laying down fixed principles, but taking care that they are to be applied according to the circumstances of each case.' For the development of the use of precedent and law reporting, see M Macnair, 'Arbitrary Chancellors and the problem of predictability' in Koops and Zwalve (n 75 above) 98; and W H D Winder, 'Precedent in Equity' (1941) 57 Law Quarterly Review 245.

105 Roodt suggests that it 'offers advisory mediation': Christa Roodt, 'State courts or ADR in Nazi-era art disputes: a choice "more apparent than real"?' (2013) 14 Cardozo Journal Conflict Resolution 421, 436. It has also been described as a 'neutral third-party facilitator', at best it is an 'innominate category': Palmer (n 65 above) 183. However, this does not take into account the status of the facilitator, not as one agreed to by parties, but as an independent, government-appointed panel. Perhaps the closest analogy can be drawn between the work of the Panel and the Financial Service Ombudsman (FSO). Thanks are due to Sarah Nield for suggesting the similarity here. Specifically, the FSO has a jurisdiction to consider claims on the basis of a test of what is fair and reasonable with a view to resolving the matter in a just manner: Financial Services and Markets Act 2000, s 228.

106 Most national museum collections (those governed by statute and in receipt of direct government funding) are prevented from transferring objects from their collections except in very limited circumstances and charities are restricted in their powers to sell objects from their core collection without appropriate authority from their regulator. Museums which are accredited under the Arts Council Scheme and which, as members of the Museums Association, are bound by its Code of Ethics are subject to ethical standards governing disposals from their collections.

in effect submits to its jurisdiction.[107] Contrastingly, where an object is in the possession of a private collection, agreement of both parties is needed, demonstrating a similarity with other methods of consensual dispute resolution, albeit in the absence of an extant legal claim.

Individualised, responsive justice

The Panel provides individualised justice, focusing on the circumstances of the claim in a similar way to Equity's original aim of responding to individual petitioners whose claims fell outside the scope of the common law writs.[108] It adopts Equity's initial dislike of excessive formalism but, unlike Equity's ultimate path, has not 'lost its useful exuberance' or 'freedom, elasticity, and luminance'.[109] Even though the Panel's starting point is the application of a broad, seemingly moralistic principle, the Panel's work has not treated claims subjectively or inconsistently. The argument is not made here that the early form of Equity administered by the Courts of Chancery is an ideal which the Panel should emulate. Instead, the Panel is conceptualised as a place in which a creative form[110] of Equity is practised which is responsive to the justice of the case. Nevertheless, there are clear parallels between the reasons for the creation of both fora and their nature as new types of claims plugging gaps left by the law.[111]

---

107 Similarly, under the Financial Services and Markets Act 2000, s 226, there is a compulsory jurisdiction for a matter to be heard by the FSO. For that reason the Spoliation Advisory Panel could be described as being, in practice, a comparable process. However, there are fundamental differences in the functions of the two processes. The FSO system was established to facilitate the quick resolution of, often relatively straightforward, claims. In most situations these would be based on legal claims that continue to exist although it is clear that the FSO can make decisions 'which do not necessarily reflect the strict legal position': *R (on the application of Bluefin Insurance Services Ltd) v Financial Services Ombudsman Ltd* [2015] Bus LR 656, 661. Contrastingly, the Panel, rather than providing swift resolution of relatively straightforward claims, provides the only forum in which claims of a moral rather than legal nature can be considered because there is no longer a legal claim to the objects. Also, the Panel has to deal with difficult evidential issues from many years earlier. For these reasons the nature of the forum is quite different and so the Panel's work justifies special consideration within the framework of Equity.

108 See Baker (n 16 above) 102–103.

109 Main (n 8 above) 384.

110 Watt describes Equity as 'a creative means to close the gap between the progress of society and the conservatism of law': Watt (n 73 above) 245. Here the creative nature of the process is based on it circumventing the injustice of applying limitation statutes to worthy cases, recognised nationally and internationally, but without undermining the need to avoid imposing retroactive laws which would interfere with property rights.

111 For a criticism of the application of equitable principles to French Holocaust claims, see Estryn et al (n 9 above) 21–51.

Furthermore, there is a close analogy with Equity when it comes to the substantive principles applied by the Panel and the remedies that it can recommend.[112]

The Panel's starting point was that of a very broad remit of dealing with claims of those who lost possession of their cultural objects during the Nazi Era. As part of this notion of individualised justice, the Panel necessarily had to deal with a varied array of factual situations, complicated by the extreme circumstances of the time, which are also affected by the passage of time which makes piecing together the evidence to construct a picture of the facts a difficult one. The Panel started with no prior cases and so had to respond to individual circumstances and determine whether these justified a conclusion that there was sufficient moral strength to the claim to, in turn, recommend a just and fair solution.[113] The Panel has received seemingly clear-cut cases of spoliation in the form of direct seizure by the Nazis,[114] although even these 'straightforward' circumstances of dispossession require careful consideration of whether the object was actually in the collection at the time of the seizure, which can be difficult to show. For example, in the *Rothberger* claim, even in the absence of categorical evidence showing the object's presence in the collection immediately before the seizure, art historical evidence showed that the collecting habits of wealthy Jewish owners meant that it was unlikely that Rothberger would have sold it.[115] In another case of seizure of a collection by the Gestapo, objects from the collection had been consigned for auction years earlier to improve the financial position of the original owner, but many objects were returned unsold, including the drawings in question and the assumption was made that given the relatively low value of the drawings they were unlikely to have been sold elsewhere prior to the seizure of the collection.[116]

The Panel has had to analyse a variety of situations involving sales of cultural objects to determine whether those sales were forced by persecution. The way in which the Panel has approached these questions

112 Support is taken here from the description of Equity (in light of the approach taken in case law) that 'equity's conscience is the interplay of an objective morality and its application to specific facts, where parties are assessed not only on what they did but on what they ought to have done': Richard Hedlund, 'The theological foundations of Equity's conscience' (2015) 4 Oxford Journal of Law and Religion 119, 139.

113 Until 2016 an additional consideration by the Panel was whether a moral obligation rested on the institution, based primarily on the actions of the museum at the time of acquisition. This now has a diminished importance (see above n 30).

114 Eg *Rothberger* claim (n 18 above) [13].

115 Ibid [10].

116 *Courtauld/Feldmann* claim and *British Museum/Feldmann* claim (n 18 above).

is dealt with below in the context of the content of claims, and the strength of the moral claims is considered in the context of decisions regarding remedies. Given its wide jurisdiction (to hear claims from anyone who *lost possession* of cultural objects *during* the Nazi Era), the Panel has also needed to determine the extent of its jurisdiction and the circumstances that amount to spoliation; it accepted that it had jurisdiction to hear a claim for an object lost during the Allied occupation of Italy in circumstances not directly attributable to the Nazis and concluded that this amounted to spoliation.[117]

In *Kreglinger v New Patagonia Meat and Cold Storage Co Ltd*[118] Lord Haldane used the metaphor of elasticity to refer to the flexibility and responsiveness of Equity's jurisdiction which was otherwise absent from an application of rigid common law rules.[119] The Panel provides a more elastic jurisdiction than the courts in procedural terms[120] and an elasticity with the scope of the claim – for it hears otherwise debarred claims on their merit. Although the Panel is not a legal tribunal tasked with hearing revitalised legal claims that have been time-barred, it provides a means of circumventing the strictures of the limitation periods. There are echoes here of the limited situations in which Equity would allow a claim even though the common law claim was time-barred. Equity can apply limitation statutes by analogy,[121] but there were some limited circumstances where Equity would depart from applying the limitation period in the case of fraud.[122] Further, although a common law action for a debt was extinguished under the common law, Equity would allow recovery of the sum owed where, for example, a testator had created a trust for the purpose of paying any debts.[123] This seems to be on the basis that, whilst the common law remedy was

117 Report of the Spoliation Advisory Panel in respect of a twelfth-century manuscript now in the possession of the British Library (23 March 2005) (2005 HC 406) (*British Library/Benevento* claim). The jurisdiction of the other four European restitution committees would not extend to these circumstances.

118 [1914] AC 25 (HL).

119 Ibid 38 and 40 (Lord Haldane). Obviously, this was in the context of the Equitable jurisdiction within the combined system.

120 Discussed in the context of providing a forum for hearing otherwise time-barred legal claims on moral grounds.

121 See Limitation Act 1980, s 36. See generally Mark Lemming, '"Not slavishly nor always" – equity and limitation statutes' in Paul S Davies, Simon Douglas and James Goudkamp (eds), *Defences in Equity* (Hart 2018).

122 Ultimately, this was reflected in the Limitation Acts, most recently in the Limitation Act 1980, s 32.

123 Eg in *Lacon v Briggs* (1844) 3 Atkyns 105. Although, as Macnair points out, the doctrine proved controversial: Mike Macnair, 'Length of time and related equitable bars 1660–1760' in Harry Dondorp, David Ibbetson and Eltjo J H Schrage (eds), *Limitation and Prescription: A Comparative Legal History* (Duncker & Humblot 2019).

barred, the right was not.[124] However, the work of the Panel goes far further than providing an alternative remedy in a situation where the right remains extant. The effect of the Limitation Acts of both 1939 and 1980 is that, where a claim is made in the tort of conversion, the right is extinguished as well as the remedy,[125] thus the Panel permits the hearing of an extra-legal claim which could result in the retransfer of a legal title which has previously been extinguished. The Panel can hear claims from anyone who lost possession of objects during the Nazi Era (where those objects are now in a national collection or other museum or gallery established for the public benefit) and the forum is available to anyone who falls within this wide category. However, the Panel restricts the circumstances in which remedies (which at its most favourable would be a re-transfer or in effect the revitalisation of that previously extinguished legal title) can be awarded by looking carefully at the moral strength of the claim to determine whether there is, in substance, a claim.

The strength of claims brought before the Panel are assessed in moral terms by focusing on the claim's *substance* (here, a claim based on circumstances of loss occasioned by systematic stripping of property as part of widespread genocide), much as the focus of Equity was on the reality of beneficial owners in trusts or on the nature of transactions which looked to be outright conveyances, but which were actually mortgages. The Panel's jurisdiction thus circumvents the *form* of the limitation statute which has extinguished both the legal right and the remedy.[126]

The Panel, in looking at the substance of the claim is required under its Terms of Reference, as part of performing its functions, to make relevant factual and legal enquiries about the cultural object with a view to assessing 'the claim as comprehensively as possible', to examine relevant evidence (assessed on the balance of probabilities) and information, to make assessments relating to the title to the object, to consider any legal restrictions and to 'give due weight to the moral strength of the claimant's case'.[127]

124 Macnair (n 123 above) 351. See also Limitation Act 1980, s 29, where, on the acknowledgment of title or payment towards a debt, the cause of action accrues again, albeit that once a right of action is barred under the Act (as it would be under section 3(2)) the cause of action cannot be revived by any subsequent acknowledgment, or payment towards a debt: s 29(7).

125 Limitation Act 1939, s 3(2) and Limitation Act 1980, s 2.

126 For, as discussed above at text to note 42, unlike most other causes of action, in the case of a conversion the Limitation Acts 1939 and 1980 extinguish both the claim and the right under section 3(2). Thus the Panel may potentially be 'conniving in the evasion of legal formalities' – a phrase used in the context of Equity: J Cartwright, 'Equity's connivance in the evasion of legal formalities' in Koops and Zwalve (n 75 above) 109.

127 SAP ToR (n 30 above) [15].

THE CONTENT AND SCOPE OF A CLAIM

The Panel's paramount purpose of seeking to achieve a just and fair solution is essentially the remedy, yet to reach that just end one first needs to analyse the scope and content of the Equitable-style claim. These matters form the focus of this part which seeks to address head-on criticisms suggesting that the 'entitlement to restitution remains unclear'[128] and that the moral claim is 'a nebulous and shape-shifting concept'.[129] Instead, it is argued that there are clear principles derived from the recommendations which show how the Panel recognises the content of a claim based on its moral strength.[130] By recognising this new form of claim the Panel is putting into practice the Equitable maxim that 'Equity will not suffer a wrong to be without a remedy'.[131]

Unlike Equity, the Panel is seeking to act on the conscience of the museum, rather than on the conscience of the original perpetrator. In only two of the 20 claims heard by the Panel has it indicated that the museum ought to have done more in the circumstances to investigate the provenance of the cultural object.[132] In only one was this clearly determinative of the final outcome.[133] The Panel's approach is,

128 De Girolamo (n 11 above) 362.

129 Ibid 381.

130 Whilst it is acknowledged that, in the case of three claims involving the same claimants, the Dutch committees have reached different conclusions from the UK Panel, nevertheless the consideration in this article is not on the inconsistency across the jurisdictions, but rather is focused on considering the internal coherence of the Panel's approach and the way in which it recognises claims and the strength of the moral claim. A discussion of those claims is outside the scope of this article (see, generally, the 'Restatement of Restitution Rules for Nazi Confiscated Art' (n 61 above)).

131 *Snell* (n 5 above) 93.

132 *British Library/Benevento* claim (n 117 above) and *Tate/Constable* claim (n 27 above). In the *Cecil Higgins/Budge* claim there appeared to be a suggestion that even where there was an inadequacy of resources, the museum was still under a moral obligation to the heirs although there was no direct criticism of the museum itself: Report of the Spoliation Advisory Panel in Respect of Four Nymphenburg porcelain figures in the possession of the Cecil Higgins Art Gallery, Bedford (20 November 2014) (2014 HC 775) [31]. In one claim the Panel indicated that further research on acquisition in the 1980s perhaps ought to have been carried out, but that the museum 'candidly concedes' that present knowledge indicated that it could have been a forced sale: Report of the Spoliation Advisory Panel in respect of three Meissen figures in the Victoria and Albert Museum (10 June 2014) (2014 HC 208) [24].

133 *British Library/Benevento* claim (n 117 above). In that claim the loss of the object, rather than being attributable to the actions of the Nazis, was due to loss occurring during the confusion of war when the Allies were in Italy. The fact that the museum had suspicions about the provenance of the object led the Panel to conclude that a moral obligation fell on the museum and for that reason ultimately recommended return of the manuscript under consideration.

therefore, overall akin to unconscionability in Equity in that there is 'an objective value judgment on *behaviour* (regardless of the state of mind of the individual in question)'[134] which necessitates action in the form of transfer of the object. Therefore, the circumstances of loss and the museum's continued retention of a cultural object of which its original owner was dispossessed means that as in Equity where 'it is appropriate to go outside the normal adversarial character of common law judicial procedure'[135] it is also appropriate to do so in the context of the Panel to give effect to its international commitments and the need to right historical wrongs.

One interpretation of the way in which the Panel has approached the claims that it hears is to interpret them as first establishing whether, *prima facie*, there is a minimum strength to the moral claim. This gives rise to a consideration of what response would represent a just and fair solution and so more nuanced factors will then determine whether the moral strength to the claim is sufficient to justify return of the object rather than one of the alternative remedies available to the Panel.[136] A minimum moral strength to a claim is established either where loss of possession was directly or indirectly at the hands of the Nazis through persecution and is thus based in the first instance on causation. There are two situations in which this arises. The first category is where the original owner was permanently deprived of their cultural object, either through force or at least direct interaction from the Nazis or their collaborators with no compensation. The second category where a minimum moral strength arises is where but for persecution the owner would not have transferred their cultural object.

An example of the first category is direct seizure – here the Panel treats such circumstances as clearly establishing a sufficient minimum moral strength to the claim. The claim has a recognisable moral strength akin to that arising where the owner of property is the victim of theft, which is reflected in legal form both in criminal and civil law.[137] The Panel seemingly accepts that there is as a clear divergence between strict legal entitlement and doing a more pleasing justice.[138] It treats the moral strength of such claims as particularly strong.[139]

134 *Cobbe v Yeoman's Row Management Ltd* [2008] 1 WLR 1752 (HL), 1788 (Lord Walker).

135 Mike Macnair, 'Equity and conscience' (2007) 27 Oxford Journal of Legal Studies 659, 681.

136 This is dealt with under the heading of 'Remedies' below.

137 Where the moral strength to the claim is so strong, the *prima facie* response is return of the object, unless legally barred from doing so.

138 This concept of Equity performing a more pleasing form of justice is set out by Watt (n 73 above) 10, discussed above at n 73.

139 *Rothberger* claim (n 18 above); *Courtauld/Feldmann* claim (n 18 above) and *British Museum/Feldmann* claim (n 18 above).

Thus, in both the *Rothberger* and *Feldmann* claims where the Gestapo seized the cultural objects in question the act of spoliation justified a response. The needs of justice are clearly far-removed from the strict entitlement[140] and so, by recognising the moral strength of the claim as sufficient for action, the Panel acts in an Equitable manner.

As De Girolamo acknowledges, 'the act of returning an item that does not belong to you, but belongs to the person who has suffered greatly by its loss, can be easily understood',[141] but she proceeds to point out that 'simplicity is not a characteristic of cultural property disputes'.[142] The Panel has, therefore, had to deal with various complicated factual situations raising nuanced moral circumstances, but it is argued that through analysis of the content of these recommendations it can be disputed that 'The most that can be gleaned is the answer – it depends',[143] and it is argued here that the Panel has developed reasoned approaches based on developing principles. This leads to the second category of circumstances raising a minimum moral strength to a claim. The Panel's basic principle here is to recognise a minimum moral strength where the original owner's transfer was caused, at least in part, by persecution by the Nazis – but this is not determinative of the extent of the moral strength to the claim, as this, as well as the appropriate way in which to respond to this to achieve a just and fair solution, depends on further factors. This means that a minimum moral strength can be established,[144] even if the sale was not at an undervalue, where the owner may have had free use of the proceeds or compensation had been paid, for these are factors that are taken into consideration when assessing the full moral strength of the claim. A determination of these then influences the decision of the most appropriate remedy to respond to the circumstances to achieve a just and fair solution for the parties.

Sales, forced by the fact of being persecuted, have been identified as satisfying the minimum threshold for establishing a strength to a moral claim. These include sales which were forced by the need to satisfy extortionate tax demands[145] or fictious debts which were levied on a person because they were a member of a persecuted group, to fund exit visas to escape further persecution, to fund the flight across Europe to escape Nazi persecution[146] or to overcome impecuniosity

140 Here the museum's legal title.

141 De Girolamo (n 11) 365.

142 Ibid.

143 Ibid.

144 Described in the *British Museum/Koch* claim as 'the minimum threshold for finding that [it] was a forced sale' (n 23 above) [19].

145 See *Glasgow City Council/attrib Chardin* claim (n 20 above) and *British Library/Biccherna* claim (n 20 above).

146 *Tate/Griffier* claim (n 22 above).

caused by fleeing that persecution.[147] By establishing one of these circumstances of loss, causation would be established. Indeed, even in the case of mixed motives for sale, the minimum threshold could be established, as was the case in the *Courtauld/Glaser* claim.[148] Here, the Panel determined that Curt Glaser sold his collection within the context of persecution (and found this to be the dominant factor for the sale) but also because of his desire to free himself of his possessions following the death of his first wife.[149] Contrastingly, where there is no causal link between Nazi persecution and a forced sale the Panel will not uphold the claim. For example, claims were unsuccessful where an object had actually been transferred by the original owner as security for a loan and was sold when the loan was called in by the bank and could not be repaid,[150] where a cultural object was sold to repay debts owed to a creditor[151] or where the sale of a cultural object took place to cover losses attributable to commercial reasons.[152] Similarly, the claim will be unsuccessful where the Panel takes the view that persecution was 'a subsidiary or causally insignificant factor' in a decision to sell a cultural object.[153]

147 *British Museum/Koch* claim (n 23 above).

148 *Courtauld/Glaser* (n 27 above).

149 Ibid [16]. Whether or not the minimum threshold for a forced sale was established in either the *Oppenheimer* claim (Report of the Spoliation Advisory Panel in Respect of an oil painting by Pierre-Auguste Renoir, 'The Coast at Cagnes', now in the possession of Bristol City Council (16 September 2015) (HC 440)) or the *Silberberg* claim (Report of the Spoliation Advisory Panel in Respect of a Gothic Relief in Ivory, now in the possession of the Ashmolean Museum, Oxford (10 February 2016) (2016 HC 777)) is less clear; in these two claims the Panel determined that any sale was forced by financial reasons, rather than by persecution, even though the original owners suffered persecution as well. In the *Oppenheimer* claim the Panel described the moral strength as being weakened by the fact that the sale was to satisfy a debt, rather than because of the persecution to which 'the Oppenheimers were undoubtedly subject' [82]. In both claims the Panel recommended that the museums display accounts of the objects' histories with the objects.

150 Report of the Spoliation Advisory Panel in respect of three Rubens paintings now in the possession of the Courtauld Institute of Art, London (28 November 2007) (2007 HC 63) [29].

151 Report of the Spoliation Advisory Panel in respect of a painting held by the Ashmolean Museum in Oxford (1 March 2006) (2006 HC 890) [35].

152 Report of the Spoliation Advisory Panel in respect of an oil sketch by Sir Peter Paul Rubens, 'The Coronation of the Virgin' now in the possession of the Samuel Courtauld Trust (15 December 2010) (2010 HC 655) (*Courtauld/Gutmann* claim), para 83. Here the Panel concluded that the sale was because of debts accrued in light of the owner's financial speculation rather than as a result of a forced sale; this conclusion was reached even though Gutmann had been arrested during the Night of the Long Knives and had suffered from loss of earnings because of Nazi anti-Semitism: [12], [73] and [75].

153 Ibid [84].

REMEDIES AND RESPONSES – 'FAIR AND JUST' SOLUTIONS

This part analyses the way in which both the nature of the Panel's remedies and the way in which it chooses the most appropriate ones to achieve 'fair and just' solutions for both parties[154] have similarities with Equitable remedies. As discussed in the previous part, the circumstances of claims brought before the Panel can differ – ranging from evidentially clear-cut seizures by the Gestapo[155] to sales of objects by owners in exile, in relative safety and in receipt of the reasonable market value with the proceeds fully at their disposal.[156] It is therefore important that the Panel has the remedial tools to respond to these different circumstances in a just and fair manner.

Creativity of remedies

Although discussions about Nazi Era looted cultural objects are frequently framed in the language of restitution (thereby focusing on the remedy rather than the claim), in the UK return is not automatic for successful claimants. Consequently, like the discretionary nature of Equitable remedies, no one particular remedy is available as of right as it would be under the common law.[157] The range of remedies that the Panel may recommend is creative in scope and can be creatively administered to best respond to the relative moral strength of the claim. The Panel can recommend return of the object, payment of compensation, an *ex gratia* payment or the display of an account of the object's history and provenance.[158] Like Equity, the Panel's remedies can focus on action and thus do a more perfect form of justice[159] rather than the, often, second-best outcome of common law damages. Monetary awards are described as ill-placed in the context of providing redress for dispossessions of culturally important (and

154 SAP ToR (n 30 above) [14].

155 Eg the *Rothberger* claim (n 18 above).

156 Eg *British Museum/Koch* claim (n 23 above).

157 Eg the Equitable discretionary remedy of specific performance depends, *inter alia*, on the inadequacy of damages, which would otherwise be available as of right for a breach of contract: *Co-Operative Insurance Society Ltd v Argyll Stores (Holdings) Ltd* [1998] AC 1, 11.

158 As well as that negotiations to implement the recommendation should be conducted as soon as possible: SAP ToR (n 30 above) [17(a)]–[(e)].

159 See Watt (n 73 above) 113 and Miller who points to Equitable remedies perfecting 'the law *interpersonally*' and '*systematically* by providing society with a set of remedies better suited to protection of important interests': Miller (n 73 above) 94 (original emphasis). In the context of Nazi Era dispossessions, return of the cultural object has been described as essential for restorative justice: T O'Donnell 'The restitution of Holocaust looted art and transitional justice: the perfect storm or the raft of Medusa?' (2011) 22 European Journal of International Law 49, 51.

financially valuable) objects within a context of systematic persecution and genocide.[160] For 'Art restitution is a painful exercise for everyone involved and requires creative thinking by all parties and a willingness to craft solutions that at first glance may appear highly unusual.'[161] Return is usually the Panel's starting point when considering remedies,[162] but alternative remedies may be recommended instead where the Panel deems the moral strength of the claim to be lower.[163]

The remedy of the display of an account of the object's history is particularly creative and allows the story of the object's wartime history to be told and the claimant's interest in the object to be acknowledged,[164] although it is unclear whether this is sufficient as a standalone remedy.[165] It reflects one aspect of restorative justice – the telling of the narrative of loss and the horrors that befell the original owners.[166] It thereby acknowledges the importance of education and remembrance of the Holocaust and Nazi crimes 'as an eternal lesson for all humanity'.[167]

Exercising a wide discretion

It has already been seen that the circumstances in which claimants lost possession of cultural objects vary significantly between the different claims heard by the Panel. These range from seizure by the Gestapo, forced sales and flight goods. When determining how to respond remedially to these situations the Panel exercises its discretion.

Discretion, as H L A Hart observes, 'occupies an intermediate place between choices dictated by purely personal or momentary whim and those which are made to give effect to clear methods of reaching clear aims or to conform to rules whose application to the particular case is obvious'.[168] In the case of the Panel, its clear aim, set out in its Terms of Reference, is to recommend appropriate action in response to a claim and to achieve a solution which is fair and just

160 O'Donnell (n 159 above) 55. See also Select Committee Seventh Report (n 17 above) vol II, Minutes of Evidence, Memorandum submitted by the Commission for Looted Art in Europe.

161 Dugot (n 33 above) 279.

162 See, for example, *Tate/Griffier* claim (n 22 above) [51].

163 See below at text to nn 198–200.

164 SAP ToR (n 30 above) [17(d)].

165 See the reservations expressed in Seventh Report of the Select Committee on Culture, *Media and Sport, 'Cultural Property: Return and Illicit Trade* (HC 1999-2000 371-II) Minutes of Evidence, Memorandum submitted by the Commission for Looted Art in Europe [47].

166 See generally O'Donnell (n 159 above).

167 Terezin Declaration on Holocaust Era Assets and Related Issues.

168 H L A Hart, 'Discretion' (2013) 127 Harvard Law Review 652, 658.

to both parties.[169] The discretion is exercised within the confines of the procedural requirements set out in the Panel's Terms of Reference, regarding the inquiries to be made, the assessment of evidence and other relevant information, the moral strength of the claimant's case and the laws that affect the respondent institution.[170] The discretion is therefore exercised following the establishment of the moral claim, discussed above and is subject to procedural matters. These represent what Dworkin describes as the 'surrounding belt of restriction' around the discretion which is left as the hole in a doughnut by that belt of restriction.[171] Whilst the term discretion can be 'easily overstated', 'it consists principally in the need to make what are sometimes fine judgments in order to apply more or less settled principles to the factual circumstances of particular cases'.[172]

The Panel's Terms of Reference set out the aim of recommending just and fair solutions, as an avowed discretion of the Panel. It is impossible to foresee every possible permutation of loss of cultural objects that the Panel will need to consider, and so therefore situations which represent borderline cases are likely to arise because it is impossible to anticipate all possible factual scenarios.

At times, exercising a discretion may include determining what is 'the fair and just thing to do or order in the instant case'.[173] It is clear that there are parallels between the court's broader powers of discretion found not only in Equity,[174] but in other statutes, when determining, for example, whether it is just and convenient to award an injunction.[175] A common theme to these various judicial pronouncements about such a discretion is that it should not alter depending on the caprice of the judge in question. In a similar vein, over time the use of the Panel's discretion to recommend just and fair solutions has developed along the lines of more settled principles which are set out below.

Simply because a variety of awards is available does not automatically mean that a jurisdiction is discretionary.[176] Instead, it is discretionary because 'more than one judicial response can legitimately be made

169 SAP ToR (n 30 above) [6] and [14].

170 Ibid [15].

171 Ronald Dworkin, *Taking Rights Seriously* (Bloomsbury 1977) 48.

172 *Snell* (n 5 above) 444.

173 Tom Bingham, *The Business of Judging: Selected Essays and Speeches* (Oxford University Press 2000) 35.

174 See *Gee v Pritchard* (1818) 2 Swanston 403; 36 ER 670, 674 (Lord Eldon LC) and Lord Blackburn's speech in *Doherty v Allman* (1878) 3 App Cas 709, 729.

175 *Beddow v Beddow* (1878) 9 Ch D 89, 93.

176 Simon Gardner, 'The remedial discretion in proprietary estoppel' (1999) 115 Law Quarterly Review 438, 443.

to any given acts'.[177] Gardner argues, in the context of proprietary estoppel, that 'It is inherent in the very idea of a discretion that the outcome is ultimately settled by men, not laws',[178] but that this can be justified providing that three conditions are satisfied. First, 'the aim of the discretion must be fixed by law';[179] secondly, 'the discretion must be necessary', and it will be necessary 'where the law properly seeks to react to multiple considerations';[180] and thirdly, 'decisions taken under the discretion must be susceptible to audit'.[181] Whilst the Panel's discretion is not fixed by law *per se*, it is set by a clear statement of policy from an international soft law instrument[182] and governed by terms of reference that have been laid before Parliament. The discretion's aim is clearly to achieve just and fair solutions for Nazi Era dispossessed owners through consideration of the moral strength of claims. Secondly, the discretion is necessary since the Panel's jurisdiction clearly seeks to react to the multiple considerations that are involved in claims regarding Nazi Era cultural objects. Thirdly, the Panel's recommendations are susceptible to audit through publication of its reports which are fully reasoned and laid before Parliament by the Secretary of State.

H L A Hart suggests that whether decisions involving discretion are rational depends on the manner in which they are made. He takes 'manner' to include not only 'narrowly procedural factors', 'the deliberate exclusion of private interest' and 'prejudice', but also 'the use of experience in the field' and also 'the determined effort to identify ... the various values which have to be considered and subjected in the course of discretion to some form of compromise or subordination'.[183] This can be seen very obviously in the case of the Panel where the Panel's recommendations draw on the experience of its varied membership to reach just and fair solutions.

A central part of the discretion is therefore the compromise between these different values and how far guiding principles assist until a point is reached where it is necessary to move beyond those guiding principles because they either do not account for the relative ignorance of facts or the relative indeterminacy of the aim.[184]

177 Simon Gardner, 'The remedial discretion in proprietary estoppel – again' (2006) 122 Law Quarterly Review 492, 504–505.
178 Ibid 502.
179 Ibid 505.
180 Ibid 507.
181 Ibid 509.
182 Washington Conference Principles (n 36 above).
183 Hart (168 above) 664.
184 Ibid 665.

Where flexibility exists in Equitable doctrines it must be exercised in a 'disciplined and principled way'.[185] It is argued here that the Panel's approach to recommending just and fair solutions for the parties (in the guise of one of the remedies of return, compensation, *ex gratia* payments or the display of an account of the object's history) is developing in a disciplined and principled way. Thus it has developed into what is described in Equity as more 'settled principles'.[186] Specifically, the Panel has sought to temper an otherwise wide discretion of 'just and fair solutions' by constraining this further than simply considering the factors set out in the Terms of Reference (discussed above in the context of establishing the moral claim).[187] This can be seen by the Panel's refusal to award 'symbolic restitution'.[188] Despite any sympathy the Panel has for the losses suffered by a claimant's family at the hands of the Nazis, the Panel's role is not to provide redress for this.[189]

The Panel has further restricted its discretion by refusing to award compensation unless claimants have continuing legal title to the claimed object.[190] In all other situations where monetary payment is appropriate it recommends *ex gratia* payments. The very nature of *ex gratia* payments is that they are paid in the absence of any legal obligation, at the beneficence of the payor or 'by favour'.[191] This approach therefore more closely reflects the moral rather than legal nature of the award. Given this moral status there is arguably an inherently wide discretion in choosing the appropriate level of award to recommend. Nevertheless, one can observe the development of settled principles applied by the Panel when quantifying *ex gratia* payments through reference to factors considered in earlier claims.[192]

The Panel has adopted a series of principles relating to the way in which it uses its discretion to achieve a just and fair solution for the

185 *Cobbe* (n 134) 1775 (Lord Walker) and *Jennings v Rice* [2003] 1 P & CR 8, 112 (Robert Walker LJ).

186 Peter Jaffey, *Private Law and Property Claims* (Hart 2007) 116.

187 The notion of decision-makers constraining their own discretion was observed by Lempert in his empirical research: Richard Lempert, 'Discretion in a behavioral perspective: the case of a public housing eviction board' in Keith Hawkins (ed), *The Uses of Discretion* (Clarendon Press 1992) 228.

188 *Tate/Constable* claim (n 27 above) [43].

189 Ibid. Instead, return will only be recommended where the circumstances of loss give a sufficiently strong moral strength to the claim (ibid).

190 *Tate/Griffier* claim (n 22 above) [53] and *British Museum/Feldmann* claim (n 18 above) [39]. Such circumstances have not yet arisen as all respondents have had prescriptive legal title; nothing in the Panel's Terms of Reference indicate that its jurisdiction is only engaged when a claimant's legal title has been extinguished – although as discussed above it is highly unlikely that legal title would have endured.

191 '*ex gratia*, adj, and adv' OED Online.

192 Woodhead (n 6 above) 190–193.

parties. To this end the Panel considers the particular circumstances of the loss and also in a bid to avoid double recompense: whether a fair value was achieved at the sale, whether the original owner had free use of the proceeds of that sale and whether appropriate compensation has already been paid for the object's loss.

The severe circumstances of loss occasioned by direct seizure have been described as 'gross acts of spoliation'[193] and in the two claims involving such circumstances, return (where legally permissible at the time) was justified.[194]

Return is the likely recommendation where the proceeds from a forced sale were less than the object's market value and where the original owner was unable to freely dispose of those proceeds;[195] these may have been placed into blocked accounts,[196] spent on exit visas or used to pay exorbitant taxes imposed on the Jewish population.[197]

Return would also be recommended where a sale was forced by the circumstances of escape from persecution and the owner had to spend those meagre proceeds on the necessities of life.[198] Contrastingly, in the *British Museum/Koch* claim the sale in the relative safety of London at a major auction house for a fair and substantial market value was treated as a forced sale at the lower end of the gravity of such sales.[199] The Panel therefore recommended the display of an account of the objects' history.[200]

In the *Courtauld/Glaser* claim, discussed above in the context of the notion of the minimum moral strength,[201] a key factor for the Panel in reaching the recommended solution of the display of an account of the object's history rather than return or a monetary response was the need to avoid double recompense. Thus, in that case both the modest compensation received by the original owner's heirs as well as the prices achieved at the auction which Dr Glaser was able to make use of all contributed to the Panel's recommendation of a commemorative remedy.[202]

---

193 *Rothberger* claim (n 18 above).

194 At the time of the claim return was not possible in the *British Museum* element of the claim, but would have been recommended had it been permissible (n 18 above).

195 Eg *Cecil Higgins/Budge* claim (n 132 above).

196 Ibid.

197 See *Glasgow City* claim (n 20 above) and *British Library/Biccherna* claim (n 20 above).

198 *Tate/Griffier* claim (n 22 above) [11]; although an *ex gratia* payment was recommended because of the Tate's then statutory bar on return.

199 *British Museum/Koch* claim (n 23 above) [25].

200 Ibid [27].

201 *Courtauld/Glaser* claim (n 27 above)[43].

202 Ibid.

Both the *Koch* and *Glaser* claims demonstrate that discretion can be used to recommend remedies that reflect the moral strength of claims which, in the case of forced sales, can be more difficult to assess than losses by seizure. However, by being able to award return in certain cases, or an account of an object's history, these provide significant scope for responding in a nuanced way to the differing moral strength of claims.

Avoiding double recompense has also been a factor when the Panel assesses the appropriate amount to award as an *ex gratia* payment. For this reason deductions have been made to the market value to reflect the costs of insurance or seller's premium that the claimants would otherwise have had to pay as well as the value of any conservation work that the museum had undertaken.[203] In a shift in practice, in one of the Panel's later claims, a claimant has been required, on return of the object, to repay the compensation that was received from the German Government after the war.[204] On the return of an object a respondent has not been required to pay anything to the claimant to reflect the public benefit derived from the cultural object whilst it was in the museum,[205] but where an *ex gratia* payment has been made an allowance to reflect the public benefit derived from the object's display in the museum has been made.[206]

Whilst the circumstances in which the original owners lost possession of the cultural object have been considered as relevant to the Panel's discretion to find a just and fair solution, the Panel has refrained from making assessments either about the uniqueness of an object or its importance to the claimant[207] when choosing remedies.[208]

What is clear, though, is that the Panel looks beyond equity between the parties. Even when the Panel is faced with the parties' preferred remedy, it will depart from this where there is no public interest in making an award of money by the taxpayer, such as where the objects are of poor quality and the public benefit to be derived from them would be low.[209] Here the more perfect form of justice for the parties (ie their preferred solution) is subordinate to the public interest.

---

203 For a general discussion of this, see Woodhead (n 6 above).

204 *Tate/Constable* claim (n 27 above) [55].

205 Ibid [60].

206 *Tate/Griffier* claim (n 22 above) [64].

207 This was compared with the relative importance of the painting to the respondent.

208 Note the Panel's reluctance to consider the public benefit of retaining the object: *British Library/Biccherna* claim (n 20 above) [32], *British Library/Benevento* claim (n 117 above) [71] and *Tate/Constable* claim (n 27 above) [46].

209 *Courtauld/Feldmann* claim (n 18 above) [28]. Instead, return was recommended.

CONCLUSION

The work of the Panel has significant parallels with the rationale for, the jurisdiction and nature of the claims and the remedies which developed in Equity. The Panel's establishment was aimed at redressing one of the gaps left by the law when dealing with Nazi Era claims and it provides a forum in which claims based on moral, rather than legal, grounds can be heard, considered and responded to. The Panel's work circumvents excessive formalism and gives effect to the substance of the claim based on broad principles. The remedies that have been recommended by the Panel have responded to the nuances of the cases, showing how it exercises its discretion to recognise circumstances with differing moral strengths in the search for just and fair solutions. Whilst on paper the Panel has a seemingly unfettered discretion to deal with claims, it has tempered this by developing principles which are akin to those found in Equity. Similarly, the Panel has developed principles to apply to deal with the substantive elements of claims that arise in a diverse range of circumstances to assess the moral strength of claims. The principles that it has adopted allow it to balance the difficult moral considerations and these approaches can be used in other claims dealing with cultural objects. By framing the claims heard by the Panel in the context of Equity in a *quasi*-legal setting, the Panel's work serves as an important model for other claims involving cultural objects in the future taken in other troubling times, for it provides legitimacy to a process that could be transposed to other situations. The recommendations are not knee-jerk reactions to claims but involve the forensic and considered treatment of historical information in an Equitable manner for all concerned.

Other contentious cultural objects lost in a variety of circumstances remain in museums with claimants having no extant legal claim and museums being unable to transfer them, even in response to a moral compunction to do so.[210] Well-known examples include the cultural objects taken from Maqdala and the Kingdom of Benin during punitive military expeditions, cultural objects taken from Aboriginal communities as well as the *cause célèbre* of the Parthenon Marbles. These situations represent similar equity gaps to the one found in the case of Nazi Era dispossessions.[211]

---

210 *AG v Trustees of British Museum* (n 48 above).

211 It is acknowledged that these different types of claims raise particular issues regarding identifying current claimant groups, the patriae to which repatriation should be made and cultural rights. Nevertheless, a forum in which to hear the claims and a process by which to assess the claims would fill the equity gap.

Frequently, strong calls for action to facilitate transfers to communities or nations from whom these objects were acquired[212] are at odds with firmly articulated arguments for retention,[213] often resulting in deadlock. The work of the Panel has shown that restitution is one of several available remedies, but that claims can be heard on moral bases in an objective manner and receive Equitable responses. Using a *quasi*-legal process applying Equitable principles as the foundation for hearing such claims has the potential to legitimise the process that is weighted neither in favour of the claimant nor the respondent.[214]

The principles used by the Panel could be adopted as a model for those other types of claims and developed accordingly. Adopting a *de facto quasi*-legal process in respect of these other claims can be justified, for, as with Nazi Era dispossessions, these same museums are in receipt of public funding and hold objects on trust for the public. These institutions are therefore similarly accountable and the circumstances vindicate comparable action to plug the gap between strict law and that 'more pleasing justice'.[215]

The focus on restraining injustice within the Equitable concept of unconscionability in these other claims is even stronger where the museums, rather than acquiring objects unaware of the gap in provenance, may have known that the objects had been obtained in campaigns of plunder or during colonial times with unequal power relations.

The creative remedies discussed above could be developed even further and might include cultural exchanges or other civil society solutions such as collaborations between museums and communities or long-term loans.[216] The development of a framework within which to exercise a *quasi*-legal discretion when seeking to achieve just and fair solutions provides an ideal model within which to assess other claims, albeit that additional categories of relevant considerations

---

212 Jeanette Greenfield, *The Return of Cultural Treasures* 3rd edn (Cambridge University Press 2007); Geoffrey Robertson, *Who Owns History?* (Biteback Publishing 2019) and Hicks (n 15 above).

213 See James Cuno, *Who Owns Antiquity? Museums and the Battle over our Ancient Heritage* (Princeton University Press 2008) and Tiffany Jenkins, *Keeping their Marbles: How the Treasures of the Past Ended Up in Museums* (Oxford University Press 2016).

214 Criticisms have been levied at situations where unequal power relations remain because repatriation decisions rest with the museums.

215 As described by Watt (n 73 above) 10.

216 Woodhead (n 1 above) 247 in the context of an application of the concept of moral title (identified in the Panel's recommendations) to other cultural heritage disputes.

may need to be added.[217] The work of the Panel therefore provides a structure and process to serve as a model for similar claims processes for other cultural heritage objects. The additional considerations as to the substance of the claim, expanded categories of remedies and the relevant considerations to take into account when recommending a remedy could populate this framework. The Equitable nature of the process and the principles it applies, as familiar and trusted ones, can serve to plug these important gaps and resolve other historical injustices involving cultural heritage objects.

Whilst Equity may not be past the age of childbearing,[218] here it has an adopted child in the form of the Spoliation Advisory Panel. It responds in an Equitable manner in the twenty-first century to claims originating over 70 years ago in circumstances beyond the comprehension of many people.

217 Eg a community's desire to allow the decay rather than preservation of it (ibid) and the relevance of the public benefit when assessing appropriate remedies.

218 *Eves v Eves* [1975] 1 WLR 1338, 1341 (Lord Denning MR) and Mark Pawlowski, 'Is Equity past the age of child bearing?' (2016) 22 Trusts and Trustees 892.

Northern Ireland
Legal Quarterly
Winter Vol. 73 No. 4 (2022) 685–716
Article DOI: 10.53386/nilq.v73i4.985

Residual liberty

Richard Edwards*

University of Exeter
Correspondence email: r.a.edwards@exeter.ac.uk

ABSTRACT

This article presents the argument that detainees do not lose their right to liberty under article 5 of the European Convention on Human Rights as currently thought. Instead, the article argues that they continue to enjoy a residual liberty which may be relied upon by detainees when challenging aspects of their detention.

Keywords: right to liberty; residual liberty; article 5 ECHR; HRA; prisoners' rights.

INTRODUCTION

When a defendant is convicted and sentenced by a trial court to a term of imprisonment it is obvious that they have lost their physical liberty because of the sentence. But have they lost *all* such liberty?[1] The answer to that simple question is not as clear-cut as might be expected. To put it another way, do prisoners retain some residual physical liberty while incarcerated? This article argues that prisoners, indeed all detainees in the custody of the state, do retain an enforceable residual liberty interest while detained. Currently, the disparate threads of authority are somewhat tangled. This means that the concept of residual liberty has not received the proper attention it deserves. This article attempts to untangle the threads of authority and weave a coherent doctrine from it. It proceeds as follows: first, we look at the origins of the problem, and how historically the law addressed the question of prisoner's liberty. Next, we consider how the issue of

* My thanks to Dr Nathan Tamblyn (Law Commission of England and Wales) for his thoughts on an earlier draft of this paper. The comments and suggestions of the anonymous referees were most helpful in improving this paper, and my thanks are also due to them. Finally, I am grateful to Lee Snook and Amelia Coughlan of Exeter's excellent Lasok Law Library for their usual unfailing and prompt assistance. The usual disclaimer, of course, applies.

1 In this article 'liberty' means the physical liberty of the individual and not a broader conception of the right based on personal autonomy as discussed, for example, by Ackermann J in *Ferreira v Levin NO* 1996 (1) SA 984 (CC), [52]–[53]. The focus on the physical liberty of detainees reflects the approach of the European Court of Human Rights under art 5 of the ECHR and the way that this guarantee has been interpreted: *Selahattin Demirtaş v Turkey (No 2)* [2020] ECHR 14305/18, [311].

residual liberty is discussed in the current case law, in three contexts: common law claims in the tort of false imprisonment; claims under the Human Rights Act 1998 (HRA); and claims before the European Court of Human Rights. Then we consider three examples of the deprivation of residual liberty: confinement in secure accommodation in a psychiatric hospital; solitary confinement – the 'prison within a prison'; and indeterminate sentences. Finally, we look to see what helpful lessons can be taken from the Canadian jurisprudence. The conclusion draws it all together.

THE EVOLVING STATUS OF PRISONERS IN ENGLISH LAW

Whilst at times English prisons have been legal black holes, English law has nonetheless intermittently protected the rights of prisoners albeit with varying degrees of rigour. In the sixteenth century James Morice observed that penal custody was 'to restrain, not to destroy; safely to guard, not sharply to punish'.[2] In a similar vein Coke, citing Bracton as authority, noted in his *Institutes* that gaolers should not inflict harm on detainees, by for example shackling them, 'because a gaol ought to be for containment and not for punishment'.[3] But more recently prisons have been an area which the judiciary have been happy to approach in 'a hands off' manner.[4] After the penal reforms of the nineteenth century the position has been that when a defendant was sentenced to imprisonment and thence transferred to prison they entered the custody of the prison governor.[5] Once convicted a felon forfeited their immediate rights and interests to the Crown, suffering a form of civil death.[6] At common law a prisoner was unable to bring and

2 Cited in Sir John Baker, *The Reinvention of Magna Carta 1216–1616* (Cambridge University Press 2017) 173.

3 Sir John Baker, 'Human rights and the rule of law in Renaissance England' [2004] 2 Northwestern Journal of International Human Rights 3, [13]. Sir Edward Coke, *Third Part of the Institutes of the Laws of England* (1797) 34.

4 James E Robertson, 'Judicial review of prison discipline in the United States and England: a comparative study of due process and natural justice' (1989) 26 American Criminal Law Review 1323, 1323–1324.

5 S 13 Prison Act 1952. Before s 58 Prison Act 1865 was enacted prisoners were at common law in the legal custody of the sheriff. See *May v Cruikshank* (1902) Cox's CC 210, 216 Wills J. The 1865 Act transferred the legal custody over prisoners to the gaoler. Parliament placed all prisons under the control of the Home Secretary with the Prisons Act 1877.

6 *May v Warden of Ferndale Institution* [2005] 3 SCR 809, [23] Le Bel and Fish JJs. Gordon E Kaiser, 'The inmate as citizen: imprisonment and the loss of civil rights in Canada' (1971) 1 Queen's Law Journal 208, 209.

maintain any legal action during their incarceration.[7] In England, even after the penal reforms of the nineteenth century, actions by prisoners contesting their treatment were unknown, for the Crown could do no wrong. But even if prisoners had been able to access a court it is unlikely that they would have enjoyed a favourable reception for the courts have long feared the ensnaring of prison administration in the 'tentacles of the law'.[8] Indeed, the courts were long content to adopt what became known as the 'hands-off' approach. Thus, in *Gibson v Young*[9] Darley CJ barred a personal injury claim by a prisoner against the Government of New South Wales on public policy grounds. Darley's approach subsequently found favour with Goddard LJ, as he then was, in *Arban v Anderson*: 'It would be fatal to all discipline in prisons if governors and warders had to perform their duty always with the fear of an action before their eyes if they in any way deviated from the [prison] rules.'[10] The fear of the chilling effect of litigation on prison administration persisted. Even in *Ex parte Germain*, a case now cited for the judgment of Shaw LJ, the majority held that, while judicial review might lie against decisions of a Board of Visitors, it would not similarly lie against the administrative decisions of prison governors.[11] Prison governors were akin to military or naval commanders, for whom disciplinary powers were an essential tool of management.[12] To allow access to the High Court would weaken the authority of prison governors and make the management of prisons very difficult. Indeed, the prospect of prison governors facing judicial review challenges was subsequently described as 'frightening' by Browne-Wilkinson LJ, as he then was.[13] Prison governors were primarily accountable to the Home Secretary whom Parliament had charged with the supervision of prisons. And in turn the Home Secretary was answerable to Parliament. Thus, as a matter of public policy the courts limited their supervisory jurisdiction when it came to applications by prisoners.[14]

---

7 A V Dicey, *Treatise on the Rule for the Selection of Parties to an Action* (Maxwell 1870) 29–30.

8 *Leech v Deputy Governor of Parkhurst* [1988] 1 AC 533, 566 Lord Bridge of Harwich.

9 (1900) 21 NSWLR 7, 12–13.

10 [1943] KB 252, 255. Lord Denning MR later noted, in a similar vein, that 'if the courts were to entertain actions by disgruntled prisoners, the governor's life would be made intolerable': *Becker v Home Office* [1972] 2 QB 407, 418.

11 *R v Board of Visitors of Hull Prison ex parte St Germain* [1979] QB 425, 447–448 (Megaw LJ) and 462–463 (Waller LJ).

12 *R v Camphill Deputy Governor ex parte King* [1984] 1 QB 735, 753. See also the similar judgments of Lawton LJ (749) and Griffith LJ (751).

13 Ibid 749 Lawton LJ.

14 Ibid 747 Lawton LJ.

Legal daylight, as Sedley termed it, has been slow to seep into this landscape.[15] Significant change began in the 1970s driven by the appointment of a less deferential judiciary against the backdrop of a more socially liberal society.[16] On the whole, these changes were to have far-reaching consequences for prison regulation. Indeed, there can be little doubt that an important turning point was the judgment of Shaw LJ in *Ex parte Germain*. In his judgment Shaw LJ set out the status of prisoners thus:

> Despite the deprivation of his general liberty,[17] *a prisoner remains invested with residuary rights appertaining to the nature and conduct of his incarceration*. Now the rights of a citizen, however circumscribed by a penal sentence or otherwise, must always be the concern of the courts unless their jurisdiction is clearly excluded by some statutory provision. The courts are in general the ultimate custodians of the rights and liberties of the subject whatever his status and however attenuated those rights and liberties may be as the result of some punitive or other process.[18]

This was significant, for here it was explicitly recognised that the rights a prisoner enjoys are attenuated or limited, but crucially not extinguished, by his or her imprisonment.[19] In other words, some rights will be unaffected by imprisonment (for example dignity) whereas others (such as liberty) will be limited and thus residual in nature. The approach of Shaw LJ has, of course, much in common with the common law principle of legality subsequently revived by the House of Lords.[20] However, unlike the principle of legality the potential of Shaw's LJ *dicta* was not fully realised, at least in English law. Instead, a subtly different approach was adopted in *Raymond v Honey* where

15 Sir Stephen Sedley, *Lions under the Throne – Essays on the History of English Public Law* (Cambridge University Press 2015) 16–18.

16 For an illuminating discussion, see David Feldman, 'Changing boundaries: crime, punishment and public law' in Jason Varuhas and Shona Wilson Stark (eds), *The Frontiers of Public Law* (Hart 2020) 281 and 290–291.

17 The liberty of the subject has long been recognised at common law. As Lord Herschell held in *Cox v Hakes* (1890) 15 AC 506, 527: 'this appeal touches closely the liberty of the subject, and the protection afforded by discharge from custody under a writ of habeas corpus. The law of this country has been very jealous of any infringement of personal liberty, and a great safeguard against it has been provided by the manner in which the Courts have exercised their jurisdiction to discharge under a writ of habeas corpus those detained unlawfully in custody.'

18 *Ex parte St Germain* (n 11 above) 455 Shaw LJ (emphasis added).

19 *R v Secretary of State for the Home Department ex parte Simms* [2000] 2 AC 115, 120, Lord Steyn observed (emphasis added) that a 'prisoner's liberty, personal autonomy, as well as his freedom of movement and association are *limited*.'

20 Ibid 120, 125–128, Lord Steyn; 131–132, Lord Hoffmann. *R v Secretary of State for the Home Department ex parte Pierson* [1998] AC 539. The phrase 'principle of legality' appears in Lord Steyn's speech at 587–589.

Lord Wilberforce held in a much cited *dicta* that 'under English law, a convicted prisoner, in spite of his imprisonment, retains all civil rights which are not taken away expressly or by necessary implication'.[21] And this presumption now represents the settled position of English law in relation to the rights of prisoners.[22] At first sight the approach of Lord Wilberforce affirms the position of Shaw LJ in offering a new approach to prisoners' rights. But the *Raymond* doctrine is neither an identical approach nor without its problems. To begin with, as Lord Jauncey of Tullichettle later observed, Lord Wilberforce fails to provide any guidance on what 'civil rights' a detained citizen is entitled to.[23] This important element of the normative framework was left undefined and would need to wait for the advent of both common law rights and the HRA to be more fully developed. At the time that *Raymond* was decided rights and freedoms were primarily residual in nature and were enforced via private law. The danger was, as later cases illustrated, that private law claims end up treating the prisoner's rights as a problem without a context. Similarly, it was unclear what standard of review would apply in determining when rights were limited. Indeed, the idea of what exactly constituted a necessary implication in this context was also unclear. And, as Richardson later noted, 'residual rights have typically been restricted by a generous interpretation of "necessary implication"'.[24] Finally, and perhaps most importantly, the Raymond doctrine does not focus adequately on the '*residuary* rights appertaining to the nature and conduct of his incarceration' which was, in fact, Shaw LJ's focus.[25] Access to court notwithstanding, the context of detention and the rights a detainee should enjoy therein, were largely left undeveloped even as the formalism of the 'hands-off' approach to penal litigation began to wane.

---

21 [1983] AC 1, 10.

22 See, for example, *R v Secretary of State for the Home Department, ex parte Leech* [1994] QB 198; *R (Daly) v Secretary of State for the Home Department* [2001] 2 AC 532.

23 *R v Deputy Governor of Parkhurst, ex parte Hague* [1992] 1 AC 58, 174. For a very useful discussion of this case, see Margaret Fordham, 'Falsely imprisoning the legally detained person – can the bounds of lawful detention ever be exceeded' (1991) Singapore Journal of Legal Studies 348.

24 Genevra Richardson, 'Prisoners and the law: beyond rights' in Christopher McCrudden and Gerald Chambers (eds), *Individual Rights and the Law in Britain* (Oxford University Press 1994), 187.

25 *Ex parte St Germain* (n 11 above), 455 Shaw LJ (emphasis added).

RESIDUAL LIBERTY AT COMMON LAW

Not surprisingly the question of residual liberty has been further considered since *Ex parte Shaw*. The approach of English law to the concept of residual liberty is exemplified by the decision in *Hague*,[26] where the Court of Appeal and House of Lords both addressed the idea. Both appeals concerned treatment within prisons, specifically further imprisonment within the prison. In essence, the appeals decided that lawful imprisonment is not made unlawful by the conditions in which a prisoner is detained within the prison. In *Hague* the applicant prisoner contended that he had been unlawfully transferred and then segregated under rule 43.[27] Hague challenged these decisions by way of judicial review and sought damages for false imprisonment caused by his unlawful segregation. Unsuccessful in the Divisional Court, Hague appealed to the Court of Appeal, which dismissed the case in judgments reminiscent of the deferential 'hands-off' approach to prisoners' rights. Before the Court of Appeal Hague successfully argued that the deputy governor had not been entitled to order his segregation in another prison, which in turn raised the question: might the deputy governor rely on the defence of lawful detention? Taylor LJ held that section 12 of the Prison Act 1952 offered a complete defence to any claim for false imprisonment. A breach of the Prison Rules could be met by the defence of lawful detention under the Prison Act.[28] Nicholls LJ was even more deferential: 'It is for the prison authorities to decide whereabouts within a prison an inmate shall be confined.'[29] Nicholls LJ continued:

> I can see no room in principle, in respect of the tort of false imprisonment, for the retention of any residual right against the prison authorities ... a prisoner's loss of freedom to go where he will is total.[30]

Indeed, it is inherent in his lawful committal to prison that a prisoner losses the ability to bring actions for false imprisonment against the prison authorities for his detention in any prison or 'any particular place within a prison'.[31] Only if prison conditions were intolerable might a prisoner be able to succeed in an action for false imprisonment.

Equally, in *Weldon* the claimant was a prisoner in Leeds prison who it was contended had been unlawfully removed from the general prison population, beaten and confined to a 'strip cell'. Weldon brought an action for false imprisonment against the Home Office

26 *Hague* (n 23 above).
27 Ibid 66.
28 Ibid 124.
29 Ibid 125.
30 Ibid.
31 Ibid.

in the County Court. It was contended on behalf of Weldon that the Prison Act 1952 required confinement in humane conditions, and in the absence of these a prisoner's detention would become unlawful. The Home Office sought unsuccessfully to have the case struck out on the grounds that a prisoner could not claim to have been deprived of any liberty by the prison authorities because he was already lawfully imprisoned. Section 12 of the Prison Act 1952, which provided the authority for the detention of prisoners, supplied a complete defence to any action for false imprisonment. An appeal against the decision of the assistant recorder to refuse to strike the claim out came before the Court of Appeal. For Ralph Gibson LJ[32] the starting point when determining what 'attenuated rights of liberty' a prisoner might enjoy was to examine the context of imprisonment, particularly the statutory framework under which a convict was imprisoned.[33] Having done this, Ralph Gibson LJ concluded that a prisoner should 'enjoy such liberty – his residual liberty – within prison as is left to him'.[34] Thus, there was no reason, His Lordship concluded, why the tort of false imprisonment should not be available to a prisoner to protect his residual liberty notwithstanding his imprisonment.[35] However, given the circumstances of the appeal Ralph Gibson LJ was reluctant to make any firm conclusions as to the merits of the claim.[36]

However, on appeal the House of Lords firmly dismissed the idea that detainees might enjoy an enforceable right to residual liberty. Indeed, the speeches of Lords Bridge and Jauncey remain the authoritative position of English law, even under the HRA. The approach of the House of Lords is unsurprisingly a paradigm of the culture of authority. While in theory prisoners are rightsholders, those rights are undefined in a positive sense and are readily attenuated in order not to frustrate the prison authorities. In keeping with the Diceyan approach to rights these are matters to be determined principally through one of statutory interpretation on the one hand, and private law remedies on the other. Thus, in the House of Lords Lord Bridge confirmed that the Prison Act provided the authority for the lawful restraint of the prisoner.[37] And therefore, while imprisoned, the confinement of the prisoner's liberty would be closely controlled by the prison authorities. Indeed, in these circumstances Lord Bridge concluded, 'the concept of the prisoner's "residual liberty" as a species of freedom of movement within the

32 His Lordship presided in the Divisional Court when it heard Hague's application. His judgment makes no reference to residual liberty.

33 *Hague* (n 23 above) 136.

34 Ibid 138.

35 Ibid 139–140.

36 Ibid 144.

37 Ibid 162.

prison enjoyed as a legal right which the prison authorities cannot lawfully restrain seems to me quite illusory'.[38] Further confinement through segregation was simply the substitution of one restraint for another. The prisoner was lawfully restrained throughout. Or to put it another way, a prisoner's liberty remains indivisible. Turning to the question of whether intolerable conditions would render the imprisonment unlawful, Lord Bridge concluded that the question would raise 'formidable difficulties' of definition. Furthermore, Lord Bridge warned, 'if the proposition be sound, the corollary must be that when the conditions of detention deteriorate to the point of intolerability, the detainee is entitled immediately to go free'.[39] As we shall see in due course this fear of inappropriate release also occurs under the HRA. Lord Bridge was joined in dismissing the idea of residual liberty by Lord Jauncey. According to Lord Jauncey placing a prisoner in segregation or a strip cell did not deprive them of any liberty which they have not already lost when initially confined. The proposition that an alteration in conditions infringed the prisoner's liberty and was thus a false imprisonment:

> presupposes that a prisoner lawfully confined in prison has, vis-a-vis the Governor, residual liberty which can be protected by private law remedies ... That a prisoner has a right to sue in respect of torts committed against him in prison is beyond doubt ... But does he have such residual liberty, vis-a-vis the governor, as amounts to a right protectable in law? I do not consider that he does.[40]

A prisoner's entire life is regulated by the prison regime stipulated by the Prison Act 1952 and the Prison Rules 1964. The confinement of a prisoner removes their liberty entirely.[41] There was no prison within the prison under English law.

Hague represents the definitive position of English law on residual liberty. However, its reasoning is both flawed and outdated. First, the definition of rights employed by the Law Lords is, of course, the residual one which was long a characteristic of English law. On that basis its conclusions are unsurprising. A prisoner enjoys a general right to liberty in the sense of the freedom left to them after the context of their imprisonment is considered. That is the approach of *Raymond*.

38 Ibid 163. Lord Ackner parted company with Lord Bridge's absolutist approach, indicating that while a prisoner would enjoy no residual liberty against the prison governor, he would nevertheless continue to enjoy it vis-à-vis other prisoners and could enforce it via the tort of false imprisonment, 166–167. Lord Ackner in the Court of Appeal gave the leading judgment in the subsequently overruled *Middleweek v Chief Constable of Merseyside* [1992] 1 AC 179.

39 *Hague* (n 23 above) 165. This is, of course, a *non sequitur*. The remedy for intolerable conditions is to require them to be made tolerable.

40 Ibid 176.

41 Ibid.

But it is neither the approach of Shaw LJ in *Ex parte Germain* nor arguably of Convention rights. Second, intolerable prison conditions can now be defined with reference to article 3 European Convention on Human Rights (ECHR).[42] But this was also true at the time. And it is anomalous that Lord Bridge was able to refer to the ECHR during *Spycatcher* but not in *Hague*.[43] And third, the consequentialist argument that an infringement of a prisoner's liberty through unlawful segregation will lead to their immediate release is a *non sequitur*, as Canadian law discussed below, amply demonstrates.

RESIDUAL LIBERTY UNDER THE HRA AND THE ECHR

Perhaps the approach of the courts to residual liberty at common law should not be entirely surprising. While the courts were more receptive to claims by prisoners, as Feldman details, they nevertheless could not entirely escape the formalism of the 'hands-off' approach.[44] However, the enactment of the HRA ought to have caused the courts to revisit the area viewing it through the lens of enforceable rights as section 6 of the Act requires. This, of course, depends in large part on the jurisprudence of the European Court, which we shall come to shortly. An opportunity to reconsider the approach of *Hague* after the commencement of the HRA arose in *Munjaz*.[45] Munjaz concerned a challenge to the legality of the policy under which patients were secluded within Ashworth Hospital. When subject to seclusion a patient would undergo

> supervised confinement and isolation ..., away from other patients, in an area from which the patient is prevented from leaving ... [on the basis that it is immediately necessary] ... for the purpose of the containment of severe behavioural disturbance which is likely to cause harm to others.[46]

In other words, the detainee is held within a secure unit within the secure unit.[47] Munjaz contended that his seclusion was unlawful on the basis that it lacked the mandatory periodic reviews required by the Mental Health Code. The hospital had adopted its own policy which failed to reflect the requirements of Code. The Court of Appeal agreed,

---

42 See, for example, *Napier v Scottish Ministers* 2005 SC 229 (the practice of 'slopping out' held incompatible with art 3 ECHR).

43 *Attorney General v Guardian Newspapers Ltd (No 1)* [1987] 1 WLR 1248, 1286.

44 Feldman (n 16 above).

45 *R (on the application of Munjaz) v Ashworth Hospital Authority* [2005] UKHL 58; [2006] 2 AC 148.

46 *Mental Health Act 1983: Code of Practice* (The Stationery Office 2015) para 26-103.

47 Brenda Hale, *Mental Health Law* 6th edn (Sweet & Maxwell 2017) para 6-028. The revised Mental Health Code expressly contemplates this form of detention.

holding that the national code ought to be followed unless there was good reason for not doing so. In the absence of such adherence there was a danger, the court concluded, that a hospital might act contrary to articles 3 and 8 ECHR. The House of Lords disagreed and dismissed Munjaz's challenge under articles 3 and 8 ECHR. However, both the Court of Appeal and the House of Lords agreed that article 5 ECHR[48] did not apply. Before the Court of Appeal, it was argued that a detainee enjoys a residual right to liberty and that as a consequence seclusion within a secure hospital, that is detention within detention, should fall within the scope of the protective ambit of article 5 ECHR. And if it was not justified under article 5(1)(e) release from seclusion should follow under article 5(4) ECHR.[49] For Hale LJ, as she then was, there was a clear division in the jurisprudence of the ECHR between the treatment of detention on the one hand and its conditions on the other. Provided that a person is detained in an appropriate institution necessary to justify the restriction on their liberty, article 5 has nothing further to say about the conditions of their detention.[50] Although it was tempting to consider further confinements within a secure institution on the Canadian idea of residual liberty, the jurisprudence of the European Court, Hale LJ concluded, did not require this.[51] Detention under article 5 was all or nothing.[52] Moreover, article 5 was procedural. Beyond ensuring that a detainee was detained in an appropriate institution, and before that determining that the original decision to detain was lawful, the jurisprudence of the European Court under article 5 was not concerned with the conditions of detention. A majority of the House of Lords agreed.[53] Lord Bingham, for instance,

48 Art 5(1) provides:

> Everyone has the right to liberty and security of person. No one shall be deprived of his liberty save in the following cases and in accordance with a procedure prescribed by law:
>
> (a) the lawful detention of a person after conviction by a competent court; ...
>
> (e) the lawful detention of persons for the prevention of the spreading of infectious diseases, of persons of unsound mind, alcoholics or drug addicts or vagrants.'

And art 5(4) further provides that 'everyone who is deprived of his liberty by arrest or detention shall be entitled to take proceedings by which the lawfulness of his detention shall be decided speedily by a court and his release ordered if the detention is not lawful.

49 *R (Munjaz) v Mersey Care National Health Service Trust* [2003] EWCA Civ 1036; [2004] 2 QB 395 [67]–[68], Hale LJ.

50 Ibid [69] Hale LJ. *Ashingdane v United Kingdom* [1985] ECHR 8225/78, [44].

51 Ibid [67]–[69] Hale LJ.

52 Ibid [70] Hale LJ.

53 Lord Bingham provided the most extensive reasons for dismissing the art 5 arguments. The remainder of the majority agreed. *Munjaz (HL)* (n 45 above) [85] Lord Hope; Lord Brown [111]; and Lord Scott agreed with Lords Bingham and Hope.

was not disconcerted that the European Court had failed to develop a concept of residual liberty.[54] Such a concept, Lord Bingham argued, would lead to not only patients challenging their seclusion but also, in the context of prisons, inmates challenging their status. Moreover, residual liberty would have the unfortunate consequence of enabling detainees to secure their unjustified release through the employment of article 5(4) ECHR.[55] Finally, while Lord Hope concluded that the conditions of a patient's detention could not be challenged under article 5 ECHR, the patient could nevertheless challenge such matters under articles 3 and 8 instead.[56]

However, the decision in *Munjaz* was not a unanimous one. On the question of whether article 5 ECHR applied Lord Steyn dissented. Invoking *Raymond*, Lord Steyn argued that a detainee continues to enjoy a residual liberty while confined. Indeed, his Lordship argued the concept was 'a logical and useful one' as was the idea of a prison within a prison.[57] Confining an individual to solitary confinement was capable of constituting 'a material deprivation of residual liberty'. Whilst *Hague* had effectively ruled this out in 1990, the enactment of the HRA now meant that this was open to question. Furthermore, Lord Steyn did not share either the scepticism of Lord Bridge that the idea of residual liberty would be employed by detainees to harass their detainer, or for that matter that private law remedies were sufficient to deal with ill-treatment meted out to detainees. For applicants like *Munjaz*, detained in secure hospitals, any unnecessary use of seclusion that involves a total deprivation of the residual liberty that they enjoy within the hospital would also amount to a further deprivation of liberty under article 5 ECHR.[58]

His domestic remedies exhausted, Munjaz, unsurprisingly, petitioned the European Court. How might the court view the idea of residual liberty? In the House of Lords, Lord Steyn had indicated that the European Court did not exclude the possibility 'that measures adopted within a prison may disclose interferences with the right to liberty'.[59] However, the settled position of the court was that disciplinary measures within prisons would not constitute deprivations of liberty.[60] As the European Court held in *Ashingdane*, provided that the initial detention had been lawfully imposed and the detaining institution

54 Ibid [30].

55 Such concerns are, of course, unfounded and rest on a confused understanding of the idea.

56 Ibid [84] Lord Hope.

57 Ibid [42] Lord Steyn. Citing the Canadian Supreme Court in *R v Miller* [1985] 2 SCR 613. Discussed below.

58 Ibid [43] Lord Steyn.

59 *Bollan v United Kingdom* [2000] ECHR 421117/98 (dec).

60 *X v Switzerland* [1977] ECHR 7754/77.

was appropriate, article 5 was not concerned with the conditions of detention.[61] Changes to such conditions were authorised by the original order authorising the detention. Indeed, this is the position of English law.[62] Notwithstanding this consistent line of authority, in *Munjaz v United Kingdom*[63] the court nevertheless interpreted article 5 ECHR in a unique manner. Indeed, the court had not interpreted article 5 ECHR in this manner before. And it has not directly done so since. But in considering the applicant's petition under article 5 ECHR the court held that where a detainee contends that has been 'a further deprivation of liberty' under that article the usual approach of the court[64] in determining whether there has been a deprivation of liberty not only applies to the further restriction but it does so with 'greater force'.[65] In other words, the court approached the matter as one of secondary (or residual) liberty but without expressly confirming this, or for that matter explaining why consideration of this further deprivation was necessary.[66] The court then went on to carefully consider whether the applicant had in fact been subject to a further deprivation of liberty when he was subject to seclusion within the secure hospital, before concluding that he had not been.[67] There was no further deprivation of liberty for four reasons. Firstly, the applicant was already detained in a high-security hospital.[68] Secondly, the applicant's seclusion was not imposed as a punishment.[69] The illness he was being treated for made him a danger to others. Thirdly, although his periods of seclusion each lasted several days, and thus tended to indicate that there was a deprivation of liberty, the court considered that factor alone was insufficient. His detention was a matter of clinical judgment by experienced practitioners.[70] And, finally, the most important factor in determining that there was no further deprivation of liberty was the fact that the seclusion regime at Ashworth was a liberal one, with the confinement balanced with association, the continual presence of staff,

61 *Ashingdane* (n 50 above) [44].

62 *R(B) v Ashworth Hospital Authority* [2005] UKHL 20; [2005] 2 AC 278, [34] Baroness Hale of Richmond. Lady Hale cited with approval the then approach of the European Court in *Ashingdane* (n 50 above).

63 [2012] ECHR 2913/06.

64 Here the court cited *Austin v United Kingdom* [2012] ECHR 39692/09, [57] which contained the long-standing authorities concerning deprivations of liberty (eg *Guzzardi v Italy* [1980] ECHR 7367/76, [92]–[93]).

65 *Munjaz* (ECHR) (n 63 above) [65]–[67].

66 Unsurprisingly, counsel for the applicant had argued his petition in part on this basis.

67 *Munjaz* (ECHR) (n 63 above) [68].

68 Ibid [69].

69 Ibid [70].

70 Ibid [71].

and meals in the ward. Seclusion at Ashworth did not amount to solitary confinement.[71] On that basis there had been no further deprivation of liberty, and consequently article 5 ECHR was not engaged. This outcome meant that the court had no need to discuss the question of remedies, particularly article 5(4) ECHR and the question of release. The Fourth Section's judgment in *Munjaz* was sadly not repeated, and the case appears not to have laid down what the Strasbourg Court describes as a 'general rule'. This is unfortunate because clearly the court considered that article 5(1) ECHR could apply where 'further deprivations of liberty', as it termed them, occur. Moreover, a divisible concept of liberty would assist the court in dealing with other article 5 ECHR cases.

After *Munjaz* a further opportunity to revisit the question of residual liberty in English law arose in *Bourgass*.[72] However, the Supreme Court eschewed the opportunity, affirming instead the pre-HRA approach of *Hague*. *Bourgass*, in essence, concerned the legality of keeping prisoners in segregation for substantial periods. Penal segregation is the paradigm 'prison within a prison'. The applicants, both serving prisoners, were segregated by order of the prison governor following episodes of violent disorder. Rule 45(1) of the Prison Rules 1999 allowed the prison governor to segregate prisoners for reasons of good order and discipline. Both prisoners were detained essentially in solitary confinement for several months until they were transferred to other prisons. Before the High Court and the Court of Appeal the applicants' case had been principally that the segregation was inherently risky, leading in some cases to suicide and permanent psychological harm.[73] However, in the Supreme Court the case was argued successfully on different grounds. Rule 45(2) allowed the prison governor to authorise the segregation of a prisoner for up to 72 hours. But segregation thereafter required the approval of the Secretary of State. As Lord Reed noted, the rationale for this further approval was simple. The governor would have the flexibility to use segregation quickly thereby effectively ensuring good order and discipline. But given the nature and dangers of segregation the continuance of such detention required the consideration of an individual independent of the day-to-day administration of the prison.[74] It followed that a segregation decision could not be taken by the governor of a prison as the delegate of the Secretary of State. The purported delegation of the segregation power under PSO 1700 was unlawful, as in turn was the segregation of the appellants. *Bourgass* is significant because the Supreme Court then

71 Ibid [72].
72 *R (Bourgass) v Secretary of State for Justice* [2015] UKSC 54; [2016] AC 384.
73 Ibid [35]–[40].
74 Ibid [86]–[89].

went on to consider whether and how the article 6(1) ECHR right to a fair trial might apply in this context. For article 6(1) ECHR to apply there must be a genuine and serious dispute over a right recognised under domestic law, and that right must be a civil one.[75] As a matter of English law, segregation was a question which was not covered by private or public law. A prisoner, on the authority of *Hague*, has neither the right to residual liberty nor any private law right to enjoy the company of other prisoners.[76] In general, the extent of association within prisons is a matter for the prison administration.[77] However, any decision to authorise the segregation of a prisoner would nevertheless be subject to ordinary judicial review principles, and this in turn would meet the requirements of article 6(1) ECHR.[78] Whether this really is the case remains open to question. Certainly, the ordinary principles of English administrative law have on occasion been held by Strasbourg to be insufficiently rigorous to protect Convention rights.[79] Only in the Court of Appeal was there a hint of a different approach. Elias LJ noted that 'whilst a prison sentence truncates [the right to personal integrity] in a major way ... it does not remove that freedom entirely'.[80] Elias LJ concluded that while 'this residual freedom of association did not enjoy any clear support in the case law of the European Court' many of the Convention rights would be engaged because of an interference with this right.[81]

THE EVOLVING JURISPRUDENCE OF THE EUROPEAN COURT

More recently, the authority of *Ashingdane* has come into question in the Strasbourg court, and this development has clear implications for our understanding of article 5 ECHR. In *Ashingdane* the European Court held that article 5 does not touch the conditions of detention. Article 5 ECHR is procedural. The original order depriving an individual of their liberty justifies subsequent deprivations or changes unless they are so remote from the basis of the original order. Detainees must, of course, be held in appropriate institutions, but beyond that article 5 ECHR was, on the authority of *Ashingdane*, silent as to the treatment

75 Ibid [106]. Art 6(1) ECHR guarantees a fair trial in the determination 'of civil rights and obligations'.

76 Ibid [122].

77 Ibid.

78 Ibid [123]–[126].

79 See, for example, *Daly* (n 22 above) [25]–[28] Lord Steyn and [32] Lord Cooke of Thornton.

80 *R(King) v Secretary of State for Justice* [2012] EWCA 376; [2012] 1 WLR 3602, [86].

81 Ibid [86]–[88].

of detainees within institutions. However, the procedural is now being transformed into the substantive as *Rooman v Belgium* shows.[82] Originally convicted in 1997 of offences including the indecent assault of a minor aged under 16 and the rape of a minor aged under 10, Rooman was a recidivist child sex offender who on the completion of his prison term in 2004 was transferred by court order to a psychiatric institution for treatment of his underlying mental ill-health. Between 2005 and 2015 the applicant applied for conditional release on three separate occasions. On each occasion the Commission de Défense Sociale declined his application on the grounds that the applicant remained a danger. Ultimately, *Rooman* petitioned the European Court arguing, amongst other things, that the failure to provide psychiatric and psychological care in his first language of German frustrated his ability to regain his liberty and was thus unlawful. The court agreed. *Rooman* is significant because in concluding that there had been a violation of article 5 ECHR the Grand Chamber announced that the time had come to 'clarify' the principles that had developed over the years in the context of the obligations of states under paragraph (1)(e) of that article. Unfortunately, the court's clarification was somewhat opaque. The court began by noting that article 5 ECHR only allows liberty to be deprived in accordance with the express provisions of the first paragraph.[83] And that an 'intrinsic link' must exist between the purpose of the deprivation and the conditions of its execution.[84] The assessment of whether conditions are suitable is to be assessed at the point they are challenged and not at the time that the detention was originally authorised.[85] Moreover, there is a positive obligation to provide treatment for those detained under article 5(1)(e) ECHR which is appropriate to their condition and assists the detainee one day regaining their liberty.[86] In fact the court could not have been clearer:

> There exists an obligation on the authorities to ensure appropriate and individualised therapy, based on the specific features of the compulsory confinement, such as the conditions of the detention regime, the treatment proposed or the duration of the detention.[87]

Crucially, the court then went on to somewhat disingenuously note that its earlier authority, principally *Ashingdane*, had always been subject to a proviso that a case could arise under article 5(1)(e) where the link between purpose of the detention and the conditions of detention

82 *Rooman v Belgium* [2019] ECHR 18052/11, [205].
83 Ibid [191].
84 Ibid [199].
85 Ibid.
86 Ibid [203]–[204].
87 Ibid [205].

were severed.[88] However, the interpretation of article 5 ECHR had evolved since *Ashingdane* so that a 'close link' was now required between the lawfulness of the detention and appropriateness of the treatment regime.[89] In other words, the lawfulness under article 5 ECHR of the detention now turned on the administration of suitable therapy for detainees.[90] The detention of the mentally ill must have a therapeutic purpose which combines an appropriate environment with real and genuine therapeutic treatment. These conditions are necessary to ensure the ultimate restoration of the detainee's liberty.[91] The court was keen nonetheless to signal that it would continue to defer to medical professionals over the exact nature of the treatment given. However, the assessment of the authorities in ensuring that the conditions of detention with a specific institution which provides appropriate treatment were nevertheless crucial. And while, of course, article 5(1)(e) could be relied upon by states to justify detention necessary for public protection this could not be used as a justification for an absence of therapeutic measures for the detainee. Thus, it would not be lawful to continue to detain an individual on the basis that it was necessary to protect the public where the detainee had been denied access to effective and appropriate treatment which would demonstrate that they were no longer a danger.[92] Finally, although the court had previously held that conditions must be primarily challenged under articles 3 and 8 ECHR it belatedly recognised that treatment might be compatible with both articles and yet 'insufficiently connected' to the purpose underlying the detention.

Whilst the decision of the European Court in *Rooman* is an important addition to our understanding of article 5, recognising that appropriate treatment is instrumental to the effectiveness of the guarantee, it is nevertheless open to criticism.[93] To begin with the court reasons primarily under the exception to the right to liberty contained in article 5(1)(e) ECHR. It does not define liberty under article 5(1) ECHR by examining the purpose of the guarantee beyond the usual broad reasons it renders almost mechanically in article 5 cases.[94] The work of understanding the nature of liberty in this context is subsumed

88 Ibid [206].
89 Ibid [208].
90 Ibid.
91 Ibid [199]–[201] and [207]–[208].
92 Ibid [210].
93 This recognition is analogous in some senses to the recognition of the European Court that the effectiveness of art 6 ECHR, the right to a fair trial, is dependent on the right of access to court. *Airey v Ireland* [1979] ECHR 6289/73, [24]. Rights do not exist in a vacuum.
94 See, for instance, *Bernstein v Bester NO* [1996] ZACC 2; 1996 (2) SA 751, [79] Ackermann J.

into the discussion of the limits to the right under article 5(1)(e). This is perhaps to some degree an understandable consequence of the institutional constitution of the court. By the time cases come before the Chamber or Grand Chamber for determination it is accepted through the admissibility stage that a right is in play. However, this leaves the definitional stage of the interpretation phase underdeveloped. Indeed, the jurisprudence would undoubtedly benefit from greater rigour at this first stage. What does 'liberty' mean? Understanding the nature and purpose of the right to liberty, indeed any right or freedom, is necessary to ensure both its effective application and what limitations on it will be upheld. But as matters stand, the court's jurisprudence currently focuses heavily on the nature of the deprivation. Thus, a richer understanding of the right to liberty eludes us because crucially the definitional aspect of its application is left underdeveloped.[95] When we turn to remedial matters the error is compounded because the court fails to grapple with the question of definition, its reasoning as to remedies is opaque at best. On one level that is understandable because the court's remedies are limited. But it does not help domestic courts, many of whom are concerned that they might need to release detainees inappropriately. A proper understanding of the nature of the right to liberty, particularly its residual nature, is vital in this context.

INDETERMINATE SENTENCING

The decision of the European Court in *Rooman* was not an entirely unexpected development as the European Court has developed similar reasoning under article 5(1)(a) ECHR with respect to the principle of rehabilitation. But this line of authority, as we shall see, is not without its difficulties and has led to disagreement with the United Kingdom (UK) Supreme Court. Both the inherent difficulties with the jurisprudence, and consequential disagreement, might have been avoided had article 5 ECHR been interpreted differently.

Under article 5(1)(a) ECHR a person may be deprived of their liberty following a conviction. Article 5(1)(a) authorises 'a penalty or other measure involving the deprivation of liberty'.[96] For this penalty to be lawful under article 5 there must be 'a sufficient causal

95 This deficiency can be seen in other contexts involving the right to liberty, for example, police powers short of a formal arrest. See Richard A Edwards, 'Police powers and article 5 ECHR: time for a new approach to the interpretation of the right to liberty' (2020) Liverpool Law Review 331, 335–337. See also the various approaches of the members of the Appellate Committee when determining whether a non-derogating control order made under s 1(2)(a) of the Prevention of Terrorism Act 2005 deprived liberty within the terms of art 5(1): *Secretary of State for the Home Department v JJ* [2007] UKHL 45; [2008] 1 AC 385.

96 *Grosskopf v Germany* [2010] ECHR 24478/03, [43].

connection between the conviction and the deprivation of liberty'.[97] The sentence of a court which deprives an offender of their liberty by way of punishment has a 'half-life'. As the sentence passes the original penological justification for the sentence, namely punishment, decays. This is particularly true in the case of dangerous offenders sentenced to indeterminate sentences, where the original reason for the detention is by its 'very nature susceptible to change' over time.[98] The link persists until it is finally broken by a supervening decision, for example a decision not to release a prisoner or to authorise their re-detention on grounds that are inconsistent with the original sentence.[99] So far, so good. But because the court has failed to reason from first principles it runs into difficulties. European jurisprudence has long held that detentions under article 5 ECHR cannot be arbitrary.[100] And for those detained under article 5(1)(a) ECHR both the sentencing order and its execution must conform with the purpose enshrined in that exception to the right.[101] In short, where a prisoner is denied an effective opportunity to work towards their rehabilitation, and thus their eventual release, the continuing detention may become arbitrary, breaking the link with the original sentence which rendered the detention lawful under article 5(1)(a) ECHR.

Over the last 15 years the European Court has firmly established the principle of rehabilitation in its jurisprudence, reflecting the progressive developments in European penal policy and practice that place greater emphasis on that principle.[102] These developments are reflected in both state practice and several of the Council of Europe's own legal instruments.[103] Rehabilitation was recognised as a tool that both addresses recidivism and fosters resocialisation. These penal goals were re-enforced by the progression principle which recognised that a prison sentence is a journey. Immediately after sentence the emphasis would naturally be on punishment and retribution. But as the sentence progressed towards completion the final stages would be

97 *Weeks v United Kingdom* [1987] ECHR 9787/82, [42]; *Kaffaris v Cyprus* [2008] ECHR 21906/04, [117].

98 *James v United Kingdom* [2012] ECHR 25119/09 57715/09 57877/09, [202]; *Weeks* (n 97 above) [46].

99 Ibid [189].

100 Ibid [192]–[195]. *Chahal v United Kingdom* [1996] ECHR 22414/93, [118].

101 Ibid [193]: Thus, in *Bouamar v Belgium* [1988] ECHR 9106/80, the European Court held that the detention of a minor under art 5(1)(d) necessitated their detention 'in an educational regime in a setting designed and with sufficient resources for that purpose'.

102 *Dickson v United Kingdom* [2007] ECHR 44362/04, [29] and [31]–[36].

103 Ibid [28]–[29]. The court also referred to art 10(3) of the International Covenant on Civil and Political Rights [29], the *UN Standard Minimum Rules for the Treatment of Prisoners* (1957) [57]–[59], and the *European Prison Rules* 1987 and 2006 [31]–[36].

on the preparation for release and, on the restoration of the offender's liberty, their reintegration into society. In *Harakchiev v Bulgaria*, for example, the court noted that, while the ECHR did not contain a right to rehabilitation, the principle was nonetheless inherent in the Convention and its guarantees.[104] All prisoners, including those sentenced to life terms, should be provided with a real opportunity to rehabilitate themselves.[105] The applicant's penal regime fell short of these standards. Harakchiev was kept in almost complete isolation, locked in his cell, and isolated from the rest of the prison population, with no social contact, work or education. In such a deleterious regime the applicant was unable to make any effective progress towards rehabilitation and a shortening of his sentence. Indeed, such were the conditions of imprisonment that the court concluded that they infringed article 3 ECHR. Earlier, in *Vinter*, the European Court was unequivocal about the importance of the rehabilitation principle: 'There is now clear support in European and international law for the principle that all prisoners ... be offered the possibility of rehabilitation and the prospect of release if that rehabilitation is achieved.'[106] In *Murray* the court further developed the principles that underpin the ECHR in this context. Firstly, every prisoner must have a reducible sentence.[107] Secondly, any detention must have legitimate penological grounds. These grounds might include punishment, deterrence, public protection and, crucially, rehabilitation. And, thirdly, the balance between these justifications is not immutable and will change over time, with different justifications being stronger at different points of the sentence. Thus, when sentenced, the interests represented by the state, such as punishment and deterrence, will be prominent. But as the sentence passes the focus will fall onto the offender, principally their rehabilitation.[108] A state may not, therefore, simply rely on the risk posed by a prisoner to justify their continued detention when they have already been imprisoned for a considerable period. In order that any such detention does not become arbitrary within the terms of article 5 ECHR a state will need to demonstrate that during the continuing detention active steps are being taken to encourage the rehabilitation of the prisoner.[109] Assessing where a prisoner may be on the penal

104 *Harakchiev v Bulgaria* [2014] ECHR 15018/11 and 61199/12, [264].

105 Ibid [265]. In the case of life prisoners, prison conditions should be conducive to the effective reform and rehabilitation of prisoners so that one day they might have their sentence reduced and eventually secure their release.

106 *Vinter v United Kingdom* [2013] ECHR 66069/09, [114].

107 Ibid [104]–[118]. That of course includes life prisoners. If a life sentence is irreducible it will be incompatible with art 3 ECHR.

108 *Murray v The Netherlands* [2016] ECHR 10511/10, [102].

109 Ibid [102].

continuum requires regular reviews and assessment undertaken against objective criteria that have a sufficient degree of clarity and certainty. Fourthly, the review must offer procedural guarantees so that at its conclusion the prisoner knows what steps they must take to secure their liberty. And finally, there must be a guarantee of effective judicial oversight.[110]

These jurisprudential developments under article 5(1), while welcome, are nonetheless problematic and neatly illustrate the problems created by the current approach to defining the right to liberty. Indeed, because the European Court fails to adequately grapple with the definition of liberty it has turned instead to the principle of legality in the form of arbitrariness to protect article 5 rights. The court recognises that article 5 guarantees a principle of rehabilitation, but not an implied right. This principle is, in effect, a positive obligation which requires states to provide an opportunity for rehabilitation by ensuring prison conditions do not jeopardise a prisoner's prospect of rehabilitation.[111] Previously, that duty arose under articles 3 and 8. But now it will also arise under article 5. However, the court has left the question of remedies in this context unclear. In the cases discussed above there was no question of release as a remedy. But if liberty has an all or nothing character then, where a court concludes there has been an infringement of article 5, surely release should automatically follow? This is a question which haunts English courts, and one which the European Court cannot adequately answer because of its approach to the definition of article 5.[112] Instead, the court resorts to a discussion of arbitrary interferences and general duties that provide little remedial assistance to national courts. On one level this is understandable, for the court enjoys a limited remedial capacity – damages and a declaration. But it remains nevertheless the authoritative body for the interpretation of the Convention. These difficulties can be seen in a series of domestic cases.

In the *Secretary of State for Justice v James*[113] the applicants were a number of prisoners sentenced to indeterminate sentences for public protection (IPPs). An IPP sentence had two parts. The first part was imprisonment for the purposes of punishment; a penal tariff. And the second part was an open-ended period of detention for public protection which followed immediately after the first part has expired. In fact, so important was the second element that Lords Brown and Judge both considered that the second element of IPP had displaced

110 Whether judicial review in England and Wales is effective for these purposes is open to question: *Vinter* (n 106 above) [109].

111 *Murray* (n 108 above) [104].

112 See, for example, *James* (ECHR) (n 98 above) [217].

113 [2009] UKHL 22; [2010] 1 AC 553.

the normal sentencing objective of rehabilitation.[114] Be that as it may, all the applicants in *James* had received short penal tariffs but had continued to be detained thereafter on public protection grounds. Implicit in the scheme was a necessity to provide IPP prisoners with rehabilitative courses so that they might be able to demonstrate to the Parole Board that they no longer posed a danger to public safety and could therefore be released. The Government failed to do so. The House of Lords decided the case on public law grounds, holding that the prisoners were entitled to rehabilitative courses. However, so far as the ECHR is concerned their Lordships decided that there could be no infringement of article 5(1) ECHR because throughout the sentence a causal link continued between the sentence of the court and the detention.[115] Although the Secretary of State was responsible for the failure to provide courses and assistance, this did not break the causal penal link. That link would only ever be broken in exceptional circumstances which were not present in this case.[116] However, James then petitioned the European Court which in turn decided that article 5(1) ECHR had been infringed.

Article 5, the European Court affirmed, was intended to prevent arbitrary deprivations of liberty. And, in general, under article 5(1)(a) ECHR the link between detention and sentence must remain unbroken. This principle applies to both the sentence and its application. Thus, although an indeterminate sentence for public protection is justified under article 5(1)(a) ECHR, it must not be applied in such a way as to amount to an arbitrary detention. Notwithstanding the state's intention to employ IPPs to protect the public, that intention could not displace the principle of rehabilitation imminent in article 5 ECHR.[117] Consequently, once the penal tariff expired and the applicant was detained on the grounds of public protection the absence of rehabilitative courses made his detention arbitrary. In effect the applicant was left unable to work effectively towards securing his liberty.[118]

James, and its difficulties, were subsequently considered by the UK Supreme Court in *Re Corey*.[119] Corey had been convicted of murder in 1973 and sentenced to life imprisonment. In 1992 he was released on licence. However, in 2010 he was recalled to prison largely because of intelligence reports that showed he had become involved

114 Ibid [48]–[49] and [100]–[101].
115 Ibid [50] and [103].
116 Ibid [51] Lord Brown; [128] Lord Judge CJ.
117 Ibid [218].
118 The court dismissed claims that the continuing detention involved interferences with arts 5(4) and 5(5) ECHR, awarding damages and a declaration.
119 *Re Corey's Application for Judicial Review* [2013] UKSC 76; [2014] 1 AC 516.

in dissident republican activity and as such posed a significant risk of harm to the public. In accordance with Northern Irish law Corey's recall was considered by a panel of parole commissioners. Before the commissioners Corey was represented by a special advocate who had access to the closed material. The commissioners decided to continue Corey's detention to protect the public from harm. Corey's subsequent challenge under article 5(4) ECHR contended that inadequate details had been provided to him in the gist, and that the decision to continue his detention was based exclusively or almost exclusively on the closed material. Lord Kerr, adopting the reasoning of the European Court in *James*, held that 'the essential question [was] whether [Corey] had an opportunity to demonstrate that the reasons that he was considered to present a threat no longer applied'.[120] The law had provided Corey with that opportunity, and for this reason his appeal was dismissed. But in doing so Lord Kerr touched upon a problem which the decision in *James* has created: where a prisoner has managed to demonstrate that their detention is no longer necessary, and thus unlawful within the terms of article 5, would their release automatically follow?[121] It was unclear, Lord Kerr observed, whether release in such circumstances was 'inevitable'.[122] Lord Mance, in a separate opinion, also canvassed the same question concluding that the European Court's decision that the applicants were denied their right to liberty was problematic. 'Logically', Lord Mance reasoned, such a finding 'implies that the prisoner should have been at once released.'[123] The European Court had failed to follow the logic of its reasoning to its natural conclusion. Furthermore, it was improbable that a prisoner denied a rehabilitative course would succeed in securing their release on that basis.[124] In reaching its conclusion that there had been a violation of article 5 the European Court made express reference to the periods where the applicant was left with access to any effective rehabilitative courses. In such circumstances, Lord Mance thought it improbable that article 5 ECHR would require prisoners to be released until the rehabilitative courses were provided, whereupon they would be recalled to prison. Moreover, the European Court had granted an award of just satisfaction for distress and frustration in *James*. This was an understandable outcome if the award was for a breach of an ancillary duty within article 5 ECHR to progress prisoners through the prison system.[125] These were not damages awarded for false imprisonment. While the

120 Ibid [48].
121 Ibid [49]–[52].
122 Ibid [52] citing *James (ECHR)* (n 98 above) [127].
123 Ibid [62].
124 Ibid [63]–[67] Lord Mance.
125 Ibid [69] Lord Mance.

idea of an ancillary duty is an attractive one, it is not one supported by authority. Moreover, it has not been taken up by the European Court in subsequent cases.

The UK Supreme Court once more returned to this question in *Kaiyam*.[126] In *Kaiyam* the applicants had all received IPP sentences, and the tariff element had expired or was about to. The applicants sought to challenge their continuing detention on Convention grounds. The principal head of challenge was a failure to provide rehabilitative courses, which it was contended had left the applicants unable to convince the Parole Board that they were no longer dangerous. In a joint judgment for the court Lord Mance and Lord Hughes declined to follow the authority and reasoning of *James*, fashioning instead an alternative approach to the application of article 5 ECHR. The UK Supreme Court was concerned, as it had been in *Corey*, that where a prisoner's detention is held to be unlawful under article 5 ECHR their release would automatically follow.[127] The approach of the European Court in *James*, Lord Mance and Lord Hughes reasoned, was supported by little if any authority.[128] *James*, their Lordships concluded, was not part of a clear and constant line of authority, and, as such, the European Court's decision did not need to be followed.[129] Instead, the court interpreted article 5 ECHR as containing an ancillary duty, which was implicit in the scheme of the article.[130] The duty was clearly not to be found in the express wording of article 5(1) ECHR. Moreover, the court concluded, it was not appropriate to derive the duty from article 5(1)(a) ECHR. Relying on article 5(1)(a) ECHR would involve employing the reasoning that the European Court had developed with respect to 'arbitrariness' and this would have 'unacceptable and implausible' consequences.[131] Consequentially, it was better to imply a duty into the overall scheme of the article. The remedy for a breach of this duty would not be release from custody, but rather an award of damages.[132] This duty, the Supreme Court concluded, would avoid the problems that they had identified as inherent in the decision in *James*: there would be no duty to release a prisoner, no fluctuation in the status of the detainees between legitimate and illegitimate detention and there was no risk to the public through the inappropriate release of dangerous offenders.[133] Kaiyam subsequently petitioned the European

126 *R (Kaiyam) v Secretary of State for Justice* [2014] UKSC 66; [2015] AC 1344.
127 Ibid [31] citing *Re Corey* (n 119 above) [63]–[69].
128 Ibid [32].
129 Ibid [18]–[20].
130 Ibid [36].
131 Ibid.
132 Ibid [39].
133 Ibid [40].

Court, arguing that the UK Supreme Court had wrongly applied *James*.[134] The European Court dismissed the petition as manifestly ill founded. In doing so, the court noted that the UK Supreme Court had declined to follow the reasoning in *James* on the basis that it would require the court to release dangerous prisoners.[135] The European Court sidestepped this disagreement, observing that how states choose to implement judgments of the court within their legal order is a matter for them.[136] The court then went on to apply the reasoning that it had developed in *James* and subsequent cases.

The issue then returned to the UK Supreme Court in the case of *Brown v The Parole Board for Scotland,*[137] where, following the decisions in *Kaiyam* and other similar cases representing a clear and consistent line of authority, the court decided that it could no longer avoid following the reasoning of *James*.[138] Lord Reed decided that this was necessary for two reasons. Firstly, he concluded that, in practice, the principle in *James* had been less demanding than the one fashioned by the Supreme Court in *Kaiyam*. Secondly, and perhaps more fundamentally, the understanding of the Supreme Court in *Kaiyam* that *James* required prisoners to be released where their lack of rehabilitative courses caused their detention within the terms of article 5 ECHR to become unlawful was wrong. The Supreme Court had overlooked the European court's part of the judgment, where the court had indicated that the availability of a remedy in the form of the Parole Board, with its discretion to release offenders who were no longer dangerous, was sufficient to comply with article 5(4) ECHR. The award of just satisfaction for a failure to provide rehabilitative courses did not necessarily entail an obligation to release the prisoner.[139] Thus, in this 'unsatisfactory situation' the UK Supreme Court had to abandon its own ancillary obligation and follow *James*.[140]

134 *Kaiyam v United Kingdom* [2016] ECHR 28160/15 (dec).
135 Ibid [71].
136 Ibid [72].
137 [2017] UKSC 69; [2018] AC 1.
138 Ibid [30].
139 Ibid [43].
140 Ibid [44].

RESIDUAL LIBERTY – THE CANADIAN DEVELOPMENT OF *EX PARTE GERMAIN*

In the common law world, the most sophisticated understanding of liberty in the context of prisons and secure units can be found in Canada. Analytically, many systems wrestle with this area.[141] The idea of residual liberty is not a creation of the Canadian Charter of Rights and Freedoms, although its development has been strengthened by it. In fact, the taproot of the idea of residual liberty lies in English law. In the leading case of *Raymond v Honey*,[142] Lord Wilberforce cited in support of his conclusion on prisoners' rights the judgment of Dickson J, as he then was, in *Solosky v The Queen*.[143] However, Dickson J was familiar with the earlier judgment of Shaw LJ having cited *Ex parte Germain* in an earlier appeal to the Supreme Court of Canada.[144] In fact, in *Martineau v Matsqui Disciplinary Board*, decided almost immediately before *Solosky*, the residual rights of prisoners had been fashioned in a way that gave proper effect to their residual right to liberty. Indeed, this case marked the beginning of a divergence between Canadian and English law based on the proper application of the reasoning in *Ex parte Germain*. In short, Canadian law has recognised that the right to liberty is divisible. This division enables Canadian law to protect both the rights of prisoners and the rule of law effectively and coherently.

In *Martineau* the Canadian Supreme Court was asked to consider whether decisions of a prison disciplinary board were subject to judicial review. Martineau had been accused of a disciplinary offence, but he contended that he had been subject to procedural irregularities that prevented him making an effective defence. The board found Martineau guilty and consequently he was sentenced to 15 days in the Special Corrections Unit (solitary confinement) on a restricted diet with a loss of privileges.[145] In finding that the board was subject to judicial review, Dickson J noted that the challenged decision:

---

141 For an interesting and useful discussion of German law, see Liora Lazarus, 'Conceptions of liberty deprivation' (2006) 69 Modern Law Review 738.

142 *Raymond* (n 21 above). The approach of Lord Wilberforce in *Raymond* is not quite the same as saying that a prisoner is invested with residual rights related to the nature and conduct of imprisonment as Canadian law shows.

143 [1980] 1 SCR 821, 839: 'A person confined to prison retains all of his civil rights, other than those expressly or impliedly taken from him by law.' Lord Wilberforce also cited *Ex parte Germain* (n 11 above).

144 *Martineau v Matsqui Disciplinary Board* [1980] 1 SCR 602, 625.

145 The full facts are set out in the judgment of Jackett CJ in the Federal Court of Canada (Appeal Division): *Martineau v Matsqui Institution* [1976] 2 FC 198, [1]–[5].

> had the effect of depriving an individual of his liberty by committing him to a 'prison within a prison'.[146] In these circumstances, elementary justice requires some procedural protection. The rule of law must run within penitentiary walls.[147]

Martineau heralded the Canadian retreat from the deferential 'hands-off' approach towards prison administration, cementing the approach of Canadian courts in their review of correctional decision making.[148] Indeed, the *Martineau* approach was soon developed in a series of cases which became known as the *Miller* trilogy.

The *Miller* trilogy, in essence, concerned the confinement of inmates to special handling units and the consequences which that might have for their residual liberty. In *Miller*, the Canadian Supreme Court recognised for the first time the idea of a 'prison within a prison'.[149] Following a disturbance at the prison where he was held, Miller was transferred to another and placed in administrative segregation in the 'Special Handling Unit'. Miller was not able to challenge the decision, or indeed to see what evidence was used to reach it. The special handling unit was reserved for dangerous offenders, who were kept segregated from the general population. Solitary confinement exemplifies the idea of 'a prison within a prison'. Such places of confinement within prisons have long existed,[150] with for example 'Little Ease' within the Tower of London being perhaps the most notorious.[151] In the companion case to *Miller*, *Cardinal*, Le Dain J noted that close confinement within a prison is 'a significantly more restrictive and severe form of detention than that experienced by the general inmate population'.[152] Confinement to the prison within a prison is a distinct and separate form of confinement that involves a significant reduction of the residual liberty of the prisoner.[153] Indeed, solitary confinement is, of itself, a

146 The term 'prison within a prison' originates from R J Sharpe, *The Law of Habeas Corpus* (Oxford University Press 1976) 149.

147 *Martineau* (n 144 above) 622.

148 *May* (n 6 above) [25] Le Bel and Fish JJs.

149 *Miller* (n 57 above).

150 *R v Shubley* [1990] 1 SCR 3, 9 Cory J.

151 R D Melville 'The use and forms of judicial torture in England and Scotland' (1905) 2 Scottish Historical Review 225, 232–233 noting that a few days in the cell 'were sufficient to break all but the stoutest spirits'.

152 *Cardinal v Director of the Kent Institution* [1985] 2 SCR 643, 653 Le Dain J. See also *R v Morin* [1985] 2 SCR 662, 671.

153 *May* (n 6 above) [28] Le Bel and Fish JJs.

severe form of punishment which can have serious consequences for the mental and physical health of the prisoner.[154]

Crucially, decisions which affect the residual liberty of a detainee, through for example extensions to the deprivation of liberty within prison, must be seen as distinct from decisions to free a prisoner from the prison system.[155] As Lamer J noted in *Dumas*:

> in the context of correctional law, there are three different deprivations of liberty: the initial deprivation of liberty, a substantial change in the nature of detention amounting to a further deprivation of liberty, and a continuation of the deprivation of liberty.[156]

Moreover, the jurisprudence has evolved to make a further distinction between primary and secondary liberty.[157] Primary liberty is restricted by a prison sentence, in other words the initial deprivation of liberty. Whereas secondary liberty is the residual liberty that remains when the primary liberty of the detainee is controlled and limited by a prison sentence. And it is this residual or secondary liberty that remains during the detention of a prisoner, preventing the state treating a detainee in an arbitrary fashion by ensuring further interferences with the prisoner's right to liberty are lawful. Confinement in a special handling unit is, as Le Dain J observed:

> a form of detention that is distinct and separate from that imposed on the general inmate population. It involves a significant reduction in the residual liberty of the inmate. It is in fact a new detention of the inmate, purporting to rest on its own foundation of legal authority.[158]

Moreover, once the applicant has demonstrated that there has been a deprivation of their residual or secondary liberty the burden of proof shifts to the defendant to justify the detention.[159] To show that the secondary deprivation of a detainee is lawful and thus consistent with their residual liberty the decision-maker would need to demonstrate that the decision was reached in a manner which was procedurally fair, evidenced and within the powers granted to the decision-maker.[160]

---

154 *Shubley* (n 150 above) 9 Cory J. Solitary confinement can border on torture. See, for example, Interim report of the Special Rapporteur of the Human Rights Council on torture and other cruel, inhuman or degrading treatment or punishment, UN General Assembly A/66/268 11-44570 2 (United Nations 2011).

155 *Miller* (n 57 above) 641 Le Dain J. *Mission Institution v Khela* [2014] 1 SCR 502, [34] Le Bel J.

156 *Dumas v Leclerc Institute* [1986] 2 SCR 469, 464 Lamer J.

157 *Idziak v Canada* [1992] 3 SCR 631, 646–647 Le Cory J.

158 *Miller* (n 57 above) 641 Le Dain J.

159 *May* (n 6 above) [71] Le Bel and Fish JJs; *Khela* (n 155 above) [40] Le Bel J. Judith Farbey, Robert Sharpe and Simon Atrill, *The Law of Habeas Corpus* 3rd edn (Oxford University Press 2011) 88.

160 *Khela* (n 155 above) [67].

But what would be an appropriate and just remedy for a prisoner whose secondary right to liberty had been infringed? The answer the Canadian Supreme Court concluded was *habeas corpus*. Thus, in *Miller*, Le Dain J held that *habeas corpus* would lie in these circumstances:

> A prisoner has the right not to be deprived unlawfully of the relative or residual liberty permitted to the general inmate population of an institution ... Any significant deprivation of that liberty, such as that effected by confinement in a special handling unit meets the first of the traditional requirements for *habeas corpus*, that it must be directed against a deprivation of liberty.[161]

In the context of a prison the writ of *habeas corpus* plays a critical role in ensuring that both the rule of law and the rights of detainees are respected.[162] *Habeas corpus* issues as of right to release a detainee from the unlawful constraint, and thus is a more powerful remedy than the discretionary ones which judicial review affords.[163] Le Dain J then went on to consider whether *habeas corpus* would only lie where it was sought to secure the unqualified liberty of the individual. Having reviewed Canadian and American authority, Le Dain J concluded that:

> in all of these cases the effect of habeas corpus is to release a person from an unlawful detention, which is the object of the remedy ... The use of habeas corpus to release a prisoner from an unlawful form of detention within a penitentiary into normal association with the general inmate population of the penitentiary is consistent with these applications of the remedy.[164]

Furthermore, it was important for *habeas corpus* to be developed and adapted to serve as an effective remedy in the context of modern incarceration.[165] Consequently, the writ was to be applied in a way that avoided a narrow and formalistic approach.[166] Of course, the writ would only lie where there was a deprivation of liberty and not a mere loss of privileges.[167] The end of the 'hands-off' doctrine did not mean that the courts would become 'hands-on'. That said, it was nevertheless crucial that *habeas corpus* be employed rather than the internal grievance procedures of institutions. Internal procedures can be protracted and are also not an appropriate forum for testing the

161 *Miller* (n 57 above) 637.

162 *May* (n 6 above) 823 Le Bel and Fish JJs.

163 *Khela* (n 155 above) [69]–[70] commenting on the necessity of avoiding a bifurcated jurisdiction that could lead to a duplication of proceedings. *Habeas corpus* was to be preferred.

164 *Miller* (n 57 above) 638.

165 Ibid 641.

166 *Khela* (n 155 above) [54].

167 *Morin* (n 152 above) 671 Le Dain J.

deprivation liberty.[168] As the court would later note, prisoners do not have the resources or the ability to discover the reasons for the change in their detention regime: 'Habeas corpus is in fact the strongest tool a prisoner has to ensure that the deprivation of his or her liberty is not unlawful.'[169]

CONCLUSION

The aim of this article has been to show that the law in this area is muddled and that greater lucidity can be afforded through the adoption of the idea of residual liberty. The disparate threads are there ready to be woven into a coherent fabric. Canadian law, using some of those English threads of authority, has already woven the idea into an elegant and principled doctrine. But in the UK, and indeed before the European Court, some limited interweaving has also begun. Thus, for example, the most obvious outcome of the dialogue between the UK Supreme Court and the European Court is that prisoners continue to enjoy the right to liberty while detained. Without addressing this point explicitly, the law has developed as a consequence of *James* to the point where a prisoner has an enforceable liberty interest which will sound in damages in appropriate cases. But quite what the extent of this article 5 ECHR liberty interest is, remains unclear. The European Court has arrived at this position via the circuitous route of arbitrary deprivations without addressing how the liberty guarantee applies in the first place. To put it another way the court has failed to reason from first principles, looking at the nature and purpose of the guarantee of the right to liberty in this context and then proceeding to see how it might apply. The *James* approach might deliver practical justice, but it does it in an incoherent manner. Indeed, the criticism of the UK Supreme Court is certainly not wide of the mark. And yet in *Munjaz* the court flirted with the idea of 'further deprivations of liberty', seemingly recognising that in detention greater scrutiny is at times required where the state wishes to further confine a detainee. This, of course, accords with the core purpose of article 5 ECHR. In *Kurt v Turkey* the European Court observed that:

> what is at stake [for detainees] is both the protection of the physical liberty of individuals as well as their personal security in a context which, in the absence of safeguards, could result in a subversion of the rule

168 *Khela* (n 155 above) [58]. Holding that the writ of *habeas corpus* ought not be declined simply because there was an alternative remedy available in the circumstances.

169 Ibid [29].

> of law and place detainees beyond the reach of the most rudimentary forms of legal protection.[170]

It is for this very reason, for example, that article 5 ECHR contains a positive obligation to ensure that, where an individual is detained, proper records are maintained.[171] The protection of articles 2, 3, and 8 ECHR are crucially dependent on this. This is, of course, equally true of detention in a prison within a prison. Placing an individual there carries with it a real and inherent risk of treatment that will engage articles 2, 3 and 8 ECHR. Perhaps more so.

However, when we turn to domestic law it offers little assistance because English law views liberty as indivisible. Had the idea of residual liberty been more fully embraced, as it has been in Canada, then arguably matters would be different. Crucially, the Canadian notion of residual liberty offers a richer understanding of liberty in the context of confinement. If we accept, as European law does, that the protection of liberty afforded by article 5 is intended to safeguard physical liberty, then, using the idea of primary and secondary liberty interests within it, we can more effectively apply that guarantee. In other words, detainees should enjoy both primary and secondary interests of liberty under the umbrella of article 5 ECHR. Detention in either a prison or secure hospital limits the primary liberty right of the individual. But the right to liberty is not extinguished by such detention. There is no negation of the essential content of the right. The right also applies to further deprivations of liberty, as the European Court suggested in *Munjaz*. A detainee continues to enjoy a secondary or residual liberty right. Where the state wishes to limit that residual liberty right, then it must proceed on the same basis that it would if the primary liberty right was at issue. That would accord with the court's view in *Munjaz*. To commit a person to solitary confinement is to imprison a prisoner within a further prison. A prisoner in such circumstances is hardly Hamlet complaining that Denmark is his prison.[172] There is a significant change in the qualitative nature of his detention. This further detention is an interference with the secondary right to liberty, but one which could be justified for disciplinary reasons. Of course, where it is not justified then article 5(4) ECHR would provide a remedy: release from solitary confinement and a transfer back into the general population of the prison. All of this would happen without touching the prisoner's primary liberty right, which throughout would remain limited by the sentence of the trial court. Similarly, a failure to provide rehabilitative courses to an IPP prisoner frustrates their ability to

170 *Kurt v Turkey* [1998] ECHR 15/1997/799/1002, [123].
171 *Anguelova v Bulgaria* [2002] ECHR 38361/97, [157].
172 William Shakespeare, *Hamlet* (Penguin 1980) act 2, scene 2 239–250, 110–111.

regain their primary liberty right. Without such courses IPP prisoners cannot demonstrate to the Parole Board that they have made sufficient progress to address their dangerousness. Where an IPP prisoner is denied access to rehabilitative courses, their secondary liberty right is engaged. Here the liberty interest is a secondary one in rehabilitation, which enables the prisoner to work towards the restoration of their primary liberty.

Naturally, a division of liberty would therefore require greater remedial flexibility. The Canadian courts have shown this is possible through the constitutional development of *habeas corpus*.[173] For example, in *Gamble*, the Canadian Supreme Court applied the writ of *habeas corpus* to protect the residual liberty interests of the appellant.[174] It was argued that *habeas corpus* would only be available to secure the complete liberty of the individual.[175] The Canadian Supreme Court disagreed. A purposive and expansive approach to the remedy of *habeas corpus* would enable the court to review the deprivation of liberty inherent in the operation of the parole ineligibility scheme without touching the sentence itself.[176] In fact, the European Court has recognised that article 5(4) may apply where it is not sought to secure the primary liberty of a prisoner, merely a change of custodial regime. Thus, in *Kuttner*[177] the applicant challenged the length of domestic proceedings where he requested a transfer from a secure hospital to an ordinary prison. In 2005 Kuttner had been sentenced to six years' imprisonment for a violent assault on his 80-year-old mother. Kuttner had a record of violent offending. Following medical evidence, Kuttner was imprisoned not in an ordinary prison but in a secure institution where his underlying mental illness could be treated. His detention was, as the European Court observed, justified under both article 5(1)(a) and (e). Between 2007 and 2009 Kuttner made a series of applications to the domestic courts seeking a transfer to the ordinary prison system before he was finally released. He subsequently complained to the European Court that these various applications had not been considered speedily as required by article 5(4). As a

173 The law has shown remedial flexibility in other contexts, eschewing the one remedy fits all violations, for example in the apply the trial within a reasonable time guarantee: *Attorney General's Reference No 2 of 2001* [2003] UKHL 68; [2004] 2 AC 72. This is another context which demands further such imaginative flexibility. Here the *lex specialis* of release contained in art 5(4) ECHR may not always be appropriate, but in those circumstances a court could grant a remedy under the *lex generalis* of art 13 ECHR. In the UK under the HRA, s 8, this would be a remedy which is 'just and appropriate'.

174 *R v Gamble* [1988] 2 SCR 595.

175 Ibid 636 Wilson J.

176 Ibid 644–646 Wilson J.

177 *Kuttner v Austria* [2015] ECHR 7997/08.

preliminary matter it was necessary for the European Court to decide if article 5(4) applied in such cases of 'parallel detention'. That is even when successful an applicant would not be released from detention, but merely transferred from one custodial regime to another. The European Court held that article 5(4) applied.[178] To decide otherwise, the court concluded, would risk rendering the applicant's detention in a secure institution immune from challenge, something which would be contrary to the object and purpose of article 5 itself. Changes to the 'category of confinement' must be open to judicial scrutiny under article 5(4) even if they did not lead to release.[179] In fact, there is no reason why a similar approach could be adopted in contexts discussed above. Just as in *Kuttner*, such remedial flexibility would avoid the spectre of releasing dangerous people while vindicating both the Convention rights of detainees and the rule of law. But, for this flexibility to be enjoyed more widely would require, as we have seen, a new approach to the right to liberty itself. The case for adoption of the idea of residual liberty has, as we have seen above, considerable force. The evolution of both English and European law in this respect is clearly a matter that requires full consideration.

178 Ibid [33].

179 [31]. *Kuttner* sits uneasily with the European Court's earlier authority such as *Ashingdane* (n 50 above), where the applicant unsuccessfully argued that art 5(4) applied in a case challenging the conditions of his detention [45], [49], and [52].

Northern Ireland
Legal Quarterly
Winter Vol. 73 No. 4 (2022) 717–741
Article DOI: 10.53386/nilq.v73i4.975

35 years later: re-examining the offence of riot in the Public Order Act 1986

Brian Cheung*
Correspondence email: brian04@gmail.com

ABSTRACT

2021 marked the 35th year of the passage of the most important public order legislation in England, the Public Order Act 1986, and saw an ambitious attempt by the Government to 'overhaul' public order law, in the form of the Police, Crime, Sentencing and Courts Bill. 2021 also marked 10 years since the devastating 2011 riots in England. In this context, this article analyses the necessity of and justifications for the riot offence. It argues that the riot offence is neither necessary from an instrumental perspective nor targeted at the mischiefs of public fear and overthrow of the state. Instead, the crux of the offence is the group element, shedding light more generally on public order law's ideological function of imposing a specific form of 'order' and its susceptibility to abuse. The riot offence should therefore be abolished.

Keywords: public order law; criminal law; riots; disorder; law reform; policing; Law Commission.

INTRODUCTION

It has been difficult to escape news of 'riots' in recent years. In July 2021, South Africa was rocked by unrest triggered by the arrest of former president Jacob Zuma. In January 2021, a mob stormed the seat of the United States (US) Congress, attempting to overturn the results of the 2020 presidential election. In May 2020, a police killing sparked mass 'Black Lives Matter' protests in the US, some of which turned violent. In 2019, Hong Kong saw months of demonstrations and street fighting between protestors and the police. Unrest has also occurred recently in the Philippines, Chile, Nigeria, amongst many other countries.[1]

Although Britain has escaped any serious recent unrest, it is only a matter of time before another major riot occurs – and not just because major riots have occurred at (almost) 10-year intervals since 1981: the Brixton riot in 1981; the poll tax riots in 1990; the Bradford and Oldham

* This article is based on the dissertation I completed whilst undertaking a Master of Laws programme at the London School of Economics and Political Science. I am grateful to Professor Conor Gearty for his guidance and supervision and to the anonymous reviewer for their comments. All views and errors are my own.

1 Karen McVeigh, 'Protests predicted to surge globally as Covid-19 drives unrest' *The Guardian* (London 17 July 2020).

riots in 2001; and the widespread riots in London and other English cities in 2011. Numerical superstitions aside, experts and leaders have warned that the Covid-19 pandemic and the Government's plan to relax conditions for police stop and search will likely lead to increased youth violence and disorder,[2] and that the conditions which led to the 2011 riots still exist.[3] Notably, at the time of writing (in 2021), there have already been at least two incidents of riot: a demonstration in Bristol in March 2021 against the Police, Crime, Sentencing and Courts Bill (the PCSC Bill) that 'turned violent', leading to eight people being charged with riot;[4] and a riot in Swansea in May 2021 for which five men were arrested on suspicion of rioting.[5]

2021 also marked 35 years since the enactment of the Public Order Act 1986 (the 1986 Act), which sets out the main public order offences in English law. The 1986 Act was enacted towards the end of a period that saw several instances of serious rioting, such as the Southall riots of 1979, the Brixton riots of 1981 and 1985 and the Broadwater Farm riot of 1985. During and after its passage, it was criticised for its failures to consider the underlying rationale for public order law as a whole[6] and to codify public order law.[7] Since then, public attention to public order issues has waxed and waned. The relative tranquillity of the last decade has meant that crime and law and order, as an issue of public interest, has taken a back seat to other political issues such as membership of the European Union, immigration and the economy.[8] But this is likely to change, with Covid-19 restrictions on gatherings, the PCSC Bill (since the time of writing enacted as the Police, Crime,

2 Jessica Murray, 'Youth violence likely to explode over summer, UK experts fear' *The Guardian* (London 23 July 2021); 'Letters: Tory crime strategy will increase risk of major public disorder' *The Guardian* (London 28 July 2021).

3 Niamh McIntyre, Pamela Duncan and Haroon Siddique, 'Conditions that led to 2011 riots still exist today, experts warn' *The Guardian* (London 30 July 2021).

4 'Bristol: eight people charged with rioting over first kill the bill protest' (*Sky News* 13 May 2021). At the time of publication, court proceedings remain ongoing: Clara Bullock, 'Woman jailed after kicking Bristol riot police' (*BBC News* 28 February 2023).

5 'Five more arrested in Mayhill riots investigation' (*ITV News* 4 June 2021). Since the time of writing, 18 people have been convicted of 'rioting' (it is unclear if all were convicted of the riot offence specifically): 'Eighteen people jailed for their part in Mayhill riot' (*South Wales Police* 19 December 2022).

6 A T H Smith, 'Law Commission Working Paper No 82 — Offences Against Public Order' [1982] Criminal Law Review 485, 486; Charles Townshend, *Making the Peace: Public Order and Public Security in Modern Britain* (Oxford University Press 1993) 161, 165.

7 A T H Smith, *The Offences Against Public Order including the Public Order Act 1986* (Sweet & Maxwell 1987) 6; Richard Card, *Public Order Law* (Jordans 2000) 3.

8 Ipsos MORI, 'Ipsos MORI Issues Index June 2020'.

Sentencing and Courts Act 2022) and recent public protests renewing awareness of public assemblies and state and police control over such assemblies.

As such, it seems appropriate now to examine the state of public order law. In this article, I focus on the most serious public order offence – riot – but much of my discussion is also relevant to public order law generally, especially the other two group offences: violent disorder and affray. I first consider the instrumental justifications for the necessity of a riot offence. I argue that the offence is unnecessary as most acts constituting riot are adequately punished by means of other offences. In the following section, I examine the main non-instrumental grounds which are said to form the bases for criminalising riot. In other words, what mischief does riot specifically target? Of the three candidates – putting the public in fear; overthrow of the state; and the so-called 'weight of numbers' – only the last one really 'fits' within the statutory definition of riot. Yet it is questionable why group behaviour or even group violence should be specifically punished. In the penultimate section, I argue that riot and other public order offences are susceptible to abuse and serve a number of objectionable ideological functions. I conclude that the offence of riot should be abolished.

IS A RIOT OFFENCE NECESSARY?

As stated in the introduction, the modern offence of riot was introduced in 1986 in the context of serious public disorder in the late 1970s and early 1980s.[9] This disorder prompted the Government to consider updating the law on public order. Following a consultation process and a report by the Law Commission[10] and a Home Office White Paper,[11] a Public Order Bill, which was largely in line with the Law Commission's recommendations,[12] was passed as the 1986 Act. The 1986 Act modernised but retained the 'principal features of the structure and application of the common law offences'.[13] The Law Commission's view was that serious offences were needed to deal with serious

9 Peter Thornton, *Public Order Law: Including the Public Order Act 1986* (Financial Training Publications 1987) 1–2. See, generally, Townshend (n 6 above) 159; Phil Scraton, '"If you want a riot, change the law": the implications of the 1985 White Paper on Public Order' (1985) 12 Journal of Law and Society 385.

10 Law Commission, *Offences Against Public Order* (Law Com Working Paper No 82 1982); Law Commission, *Offences Relating to Public Order* (Law Com No 123 1983).

11 Home Office and Scottish Office, R*eview of Public Order Law* (Cmnd 9510 1985).

12 Richard Card, *Public Order: The New Law* (Butterworths 1987) 11.

13 Law Commission, *Offences Relating to Public Order* (n 10 above) para 2.2.

disturbances to public order and that changes should be made with caution in an area of law closely connected with individual liberties.[14]

The 1986 Act replaced the common law offence of riot with a new statutory offence:

> Where 12 or more persons who are present together use or threaten unlawful violence for a common purpose and the conduct of them (taken together) is such as would cause a person of reasonable firmness present at the scene to fear for his personal safety, each of the persons using unlawful violence for the common purpose is guilty of riot.[15]

The statutory offence differs from the common law offence in a few ways, including that the minimum number of persons required is increased from 3 to 12 and that the defendant must actually use violence rather than just threaten it.[16] However, as intended by the Law Commission, the principal features of the common law offence remain, such as the need for group violence, a common purpose and a hypothetical bystander to fear for their personal safety (albeit that at common law there was some doubt as to whether an actual bystander needed to be present).[17]

During the consultation period, some commentators made submissions to the Law Commission that the riot offence should be abolished without replacement.[18] The National Council for Civil Liberties, for example, argued that rioters could be charged with 'simpler offences ... with adequate maximum penalties'.[19] The use of violence required to commit an offence of riot will almost always fall under one or more 'mainstream' offences,[20] such as criminal damage (including arson), assault, possession of an offensive weapon, public nuisance, theft and burglary. It is noteworthy that a number of these overlapping offences carry significant penalties at least as heavy as the maximum sentence of 10 years' imprisonment for riot; criminal damage also carries a maximum sentence of 10 years' imprisonment,[21] whilst aggravated criminal damage and arson each carry a maximum sentence of life imprisonment.[22] Even burglary, when committed in respect of a dwelling, has a higher maximum sentence than riot: 14

14 Ibid.
15 Public Order Act 1986, s 1.
16 Thornton (n 9 above) 9.
17 Ibid 59.
18 Law Commission, *Offences Relating to Public Order* (n 10 above) para 1.12.
19 Peter Thornton, *We Protest: The Public Order Debate* (National Council for Civil Liberties 1985) 23. See also Card (n 12 above) 11–12; Card (n 7 above) 83.
20 Thornton (n 19 above) 24–25; Smith (n 7 above) 76–77; Law Commission, *Offences Against Public Order* (n 10 above).
21 Criminal Damage Act 1971, s 4.
22 Ibid.

years,[23] which may surprise laypersons, who may think that the public order offence of riot is more serious than a property offence of burglary.[24]

The statistics on riot convictions in Table 1 and Table 2 show that the riot offence has been used, as envisaged by the Law Commission and the Home Office,[25] primarily to deal with the most serious situations of disorder. The general prevalence of riot convictions is low, with the number of offenders convicted or cautioned between 1987 and 2009 (inclusive) averaging 17 and the number of defendants convicted between 2004 and 2020 (inclusive) averaging just 4.5. The latter period includes the timeframe between 2014 and 2020 (inclusive) when no one was tried for riot. There are notable spikes in the numbers for 2002 and 2011, which are likely to relate to the Bradford and Oldham

Table 1: Number of offenders cautioned or convicted for riot between 1987 and 2009 (inclusive)[26]

Year	Convicted or cautioned
1987	8
1988	30
1989	30
1990	3
1991	10
1992	31
1993	18
1994	3
1995	11
1996	11
1997	0
1998	0
1999	0
2000	2
2001	10
2002	137
2003	46
2004	13
2005	7
2006	1
2007	3
2008	5
2009	9

23 Theft Act 1968, s 9(3)(a).

24 Carly Lightowlers and Hannah Quirk, 'The 2011 English "riots": prosecutorial zeal and judicial abandon' (2015) 55 British Journal of Criminology 65, 71.

25 Home Office and Scottish Office (n 11) para 3.16; Law Commission, *Offences Relating to Public Order* (n 10 above) paras 2.10–2.11.

26 Data sourced from Ministry of Justice, Criminal Statistics: England and Wales 2009 Statistics Bulletin: Annex A: Additional Tables' (October 2010) (Table 11); Home Office, *Criminal Statistics: England and Wales 2000: Statistics Relating to Crime and Criminal Proceedings for the Year 2000* (Cm 5312 2001) 126 (Table 5.18); Home Office, *Criminal Statistics: England and Wales 1997: Statistics Relating to Crime and Criminal Proceedings for the Year 1997* (Cm 4162 1998) 52 (Table 5.18).

Table 2: Number of defendants tried for riot between 2004 and 2020 (inclusive)[27]

Year	Tried	Convicted	Acquitted
2004	13	12	1
2005	14	7	7
2006	1	1	0
2007	5	3	2
2008	3	3	0
2009	12	9	3
2010	1	0	1
2011	1	1	0
2012	27	20	7
2013	1	1	0
2014	0	0	0
2015	0	0	0
2016	0	0	0
2017	0	0	0
2018	0	0	0
2019	0	0	0
2020	0	0	0
Total	**78**	**57**	**21**

riots of 2001 and the English riots of 2011 respectively. Although these statistics suggest that the riot offence remains a tool used by prosecutors in practice, they do not shed any light on whether alternative offences could have been charged.

However, the data specifically relating to the 2011 riots do suggest that the riot offence is significantly less used compared to other offences. The figures for police recorded crime during the 2011 riots show that only 3 per cent of recorded crimes fell under 'disorder', with only 1 per cent being 'violent disorder' (the riot offence is not specifically documented).[28] By contrast, acquisitive crimes such as

27 Data sourced from Ministry of Justice, 'Criminal justice system statistics publication: Crown Court: pivot table analytical tool for England and Wales (12 months ending December 2004 to 12 months ending December 2014)' (21 May 2015) and Ministry of Justice, 'Criminal justice system statistics publication: Crown Court: pivot table analytical tool for England and Wales (12 months ending December 2010 to 12 months ending December 2020)' (20 May 2021).

28 Home Office, 'Overview of recorded crimes and arrests resulting from disorder events in August 2011' (October 2011) 12.

burglary comprised half of all recorded crimes and criminal damage comprised 36 per cent.[29] The data on convictions are even more telling: there were only 21 defendants convicted of riot in 2011 and 2012,[30] compared to 2158 defendants proceeded against at magistrates' courts for all offences (as of 10 August 2012) that related to the 2011 riots.[31] Riot therefore accounted for, at most, 1 per cent of the offences for which participants of the 2011 riots were convicted.

The Law Commission took a different view on the need for a riot offence.[32] One of its conclusions was that, without the offence, 'the law would not be able to deal adequately with those who provoke or lead wide-scale public disturbances and who resist the efforts of the police to restore order'.[33] But this reasoning is questionable: in terms of the severity of punishment, we have already discussed equally heavy sentences being available for the more 'mainstream' offences that are likely to cover acts constituting riot.

As for provocation or leadership, this seems to be a narrow mischief that could be addressed with a more specific offence than riot. The typical case of a riot offender will be a participant, rather than a leader, of a riot. In any event, provocation or leadership would likely fall under the inchoate offences of either encouraging or assisting an offence (commonly known as incitement, after the common law offence that it replaced) or conspiracy.

Of course, it may be that, where a defendant has encouraged a riot in general terms rather than encouraging any specific form of violence, a charge of incitement to riot is an easier route for the prosecution than having to prove incitement to another substantive offence. A case in point is provided by the convictions of two men for 'inciting rioting' on Facebook during the 2011 riots.[34] The specific offence was encouraging or assisting offences believing that one or more would be committed, contrary to section 46 of the Serious Crime Act 2007. The prosecution established that one of the defendants, Blackshaw, had encouraged riot, burglary and criminal damage.[35] He was sentenced to four years' imprisonment,[36] a sentence that could have been imposed even without the riot element of the charges. However, the prosecution was unable or unwilling to prove that the other defendant,

29 Ibid.
30 See Table 2.
31 Ministry of Justice, 'Public disorder of 6th–9th August 2011 statistical tables – September 2012' (13 September 2012) Table 1.2.
32 Law Commission, *Offences Relating to Public Order* (n 10 above) paras 6.7–6.10.
33 Ibid 6.8.
34 *R v Blackshaw* [2011] EWCA Crim 2312.
35 Ibid [54].
36 Ibid.

Sutcliffe, had encouraged any offence other than riot.[37] Sutcliffe had created an event invitation on Facebook to meet in Warrington for 'The Warrington Riots', which included a photograph of riot police in a 'stand off position' with a group of 'rioters'.[38] He eventually realised the enormity of his action and cancelled the event, although there is doubt as to whether he did this knowing that the police were searching for him.[39] In any event, no riot occurred in Warrington.[40]

Setting aside the lack of actual harm and the question of whether inciting an offence that does not subsequently occur should be criminalised, on which much academic ink has been spilled,[41] the choice of charge (incitement to riot rather than incitement to an offence against the person or property) suggests that the prosecution was unable even to identify specific persons or property that could have been *potentially* harmed. In these circumstances, it is questionable whether Sutcliffe deserved any punishment and, if so, why. I return to this issue in the next section, when I discuss the mischief that riot tackles.

Even if Sutcliffe deserved punishment, alternatives to riot were available. For example, he could conceivably have been arrested and bound over,[42] albeit that this would not have resulted in a conviction. The charge could also have been public nuisance instead,[43] which has been expressly recognised by the House of Lords as overlapping with public order offences.[44] Public nuisance is:

> doing an act not warranted by law, or omitting to discharge a legal duty, where the effect of the act or omission is to endanger the life, health, property or comfort of the public, or to obstruct the public in the exercise of rights common to everyone.[45]

37 Ibid [59].
38 Ibid [60].
39 Ibid [71].
40 Ibid [61]–[63].
41 See eg Andrew Ashworth, 'Defining criminal offences without harm' in Peter Smith (ed), *Criminal Law: Essays in Honour of J C Smith* (Butterworths 1987); Celia Wells and Oliver Quick, *Lacey, Wells and Quick: Reconstructing Criminal Law: Text and Materials* 4th edn (Cambridge University Press 2010) 311–316.
42 *Lansbury v Riley* [1914] 3 KB 229 (KB).
43 Cf *R v Madden* (1975) 1 WLR 1379 (CA) where the court accepted that a bomb hoax telephone call could constitute a public nuisance if it affected a sufficiently wide class of the public.
44 *R v Rimmington* [2005] UKHL 63.
45 Ibid.

Public nuisance is problematic in terms of being even more broadly defined than riot.[46] However, if Sutcliffe was to be punished at all, public nuisance would have 'fit' better given that the consequences caused were not an actual riot or any actual harm but people being 'appalled', 'put in fear' and 'disturbed'.[47] In fact, making hoax bomb calls and making a video threatening to bomb an aircraft have been successfully prosecuted using public nuisance charges.[48] The statutory replacement for public nuisance proposed in the PCSC Bill makes this 'fit' even clearer, as it targets conduct that causes 'serious distress, serious annoyance, serious inconvenience and serious loss of amenity'.[49]

The offence of riot may have a further instrumental function of enabling a defendant to be charged where there are evidential difficulties preventing a charge of a specific offence of violence to the person or property: that is, where they can be proved to have participated in a group which has harmed persons or property but cannot be proved to have actually committed that harm themselves.[50] This argument was made by the Law Commission in its Working Paper,[51] but was absent in its final report.[52] Perhaps the Law Commission recognised what Thornton calls the 'danger of using … "a crime of the utmost importance in the law of public order" to circumvent "evidential difficulties" in proving guilt',[53] although it did continue to rely on a similar argument in relation to affray.[54] In any event, in such circumstances, a charge of joint enterprise to commit one of the offences against the person or property could be brought against all the members of the group, provided the necessary intent is proven.[55]

There is one more gap that the riot offence might fill. Riot requires the offender to have used violence, but 'violence' is defined broadly (and circularly) in the 1986 Act: it is 'any violent conduct' and there

46 The House of Lords and the Law Commission accepted that public nuisance is sufficiently certain: ibid; Law Commission, *Simplification of Criminal Law: Public Nuisance and Outraging Public Decency* (Law Com Consultation Paper No 193 2010) paras 4.2–4.7.

47 *R v Blackshaw* (n 34 above) [72].

48 Law Commission, *Simplification of Criminal Law: Public Nuisance and Outraging Public Decency* (Law Com No 358 2015) para 3.16.

49 Police, Crime, Sentencing and Courts HL Bill (2021-22) 40, cl 60(2)(c). Since the time of writing, this clause has been enacted as s 78 of the Police, Crime, Sentencing and Courts Act 2022.

50 Thornton (n 19 above) 24.

51 Law Commission, *Offences Against Public Order* (n 10 above) para 5.11.

52 Law Commission, *Offences Relating to Public Order* (n 10 above) paras 6.7–6.10.

53 Thornton (n 19 above) 24.

54 Law Commission, *Offences Relating to Public Order* (n 10 above) para 3.5.

55 *R v Jogee* [2016] UKSC 8 [1], [88]. See also Thornton (n 19 above) 24–25.

is no need for any person or property to be actually harmed or for any intent to do such harm. The violence merely needs to have been capable of causing such harm.[56] In theory, therefore, riot may cover acts of 'violence' that do not fall under any of the mainstream offences (except public nuisance, which is so broad that it arguably could cover all acts capable of constituting riot), many of which require actual harm. In practice, however, it is difficult to think of many such acts. Thornton suggests that 'violent' conduct could include 'running through the streets in a gang, heavy pushing at barriers or police cordons, and even a large number of pickets or demonstrators shouting threats to dissuade people from continuing to work'.[57]

Whilst some people would characterise these actions as 'violent', I would suggest that many would not. In any event, the point is that 'violence' for the purposes of the riot offence is so vague that it is unlikely to meet the 'standard of clarity and precision' that Lord Sumption held is required of the elements of a criminal offence.[58] This is particularly important in the context of public order law, which implicates fundamental rights and principles such as the right to freedom of assembly and equality before the law and can restrict the civic life that is crucial to the flourishing of democracy. Furthermore, imprecision enables abuse. I return to this theme later in this article. For now, the looseness of the definition of 'violence' leads to the question, 'Why should the violence that constitutes riot be specifically punished?'

WHAT MISCHIEF DOES RIOT TARGET?

In the previous section, I argued that riot overlaps with offences against the person and property in that acts constituting riot would almost always fall under one of the latter offences. However, I conceded that certain acts of 'violence' that can cause harm, but which do not in fact cause harm, can constitute riot but not fall under one of the 'mainstream' offences. This raises the question of why these acts should be criminalised under riot.

Furthermore, it might be argued that, even where acts constituting riot could be punishable as an offence against the person or property, there is something different about those acts, morally or otherwise, that makes riot the most appropriate offence to convey censure. This argument, which is essentially about labelling,[59] is distinct from the

56 1986 Act, s 8.

57 Thornton (n 9 above) 11.

58 *R (Nicklinson) v Ministry of Justice* [2014] UKSC 38, [239].

59 See, by way of analogy, the Law Commission's arguments in relation to the need for a public nuisance offence: Law Commission, *Simplification of Criminal Law: Public Nuisance and Outraging Public Decency* (n 48 above) paras 3.23–3.26.

instrumental arguments I discussed in the last section, which related to the need to secure punishment in the sense of 'hard treatment'. The Law Commission's final report alludes to this, arguing that the absence of a riot offence would '[fail] to give significant recognition to the factors of the weight of numbers used for a common purpose, to which ... considerable importance is attached by the common law'.[60]

What, then, is different about using violence as part of a large group of people with a common purpose, compared to the use of violence by an individual, and violence that falls short of the 'mainstream' offences? To answer this question, I consider the larger question of what social values and interests public order law seeks to protect and preserve. Smith identifies two candidates: one is the 'inchoate' interest of preventing public fear of physical harm being caused to persons or property,[61] and the other is nothing less than the constitutional stability of the country.[62] I will discuss the latter first.

It is true that some riots were perceived historically as threatening the rule of the reigning monarch or the state, particularly during the Hanoverian era.[63] However, the common law offence of riot was a misdemeanour only,[64] and there is some authority that it was limited to riots with a private common purpose.[65] Hence, the offence was not directed at riots with the character of rebellions and insurrections, but at the less threatening riots that were commonplace from the late seventeenth century to the early nineteenth century.[66] During this period, riots tended not to be seen as threats to the state,[67] and the English elite 'lived on rather casual terms with popular volatility as long as the latter did not ... challenge the fundamentals of the current system'.[68] Riots were not concerned with overthrowing the social order

60 Law Commission, *Offences Relating to Public Order* (n 10 above) para 6.9.

61 Smith (n 6 above) 486; Smith (n 7 above) 1–2.

62 Smith (n 7 above) 1–3.

63 Adrian Randall, *Riotous Assemblies: Popular Protest in Hanoverian England* (Oxford University Press 2006) 23–24.

64 Smith (n 7 above) 2; Michael Supperstone, *Brownlie's Law of Public Order and National Security* 2nd edn (Butterworths 1981) 120.

65 Law Commission, *Offences Relating to Public Order* (n 10 above) para 6.25; Supperstone (n 64 above) 132; Card (n 12 above) 18.

66 David Williams, *Keeping the Peace: The Police and Public Order* (Hutchinson 1967) 12; Carl J Griffin, *Protest, Politics and Work in Rural England, 1700–1850* (Palgrave Macmillan 2014) xiii.

67 Townshend (n 6 above) 10–11.

68 Allan Silver, 'The demand for order in civil society: a review of some themes in the history of urban crime, police, and riot' in David J Bordua (ed), *The Police: Six Sociological Essays* (John Wiley & Sons 1967) 19. See also Randall (n 63 above) 20; Townshend (n 6 above) 10.

but were often a form of protest or collective bargaining, with specific demands that could be accommodated within the existing order.[69]

When a riot came to be perceived as a threat to the state's authority, it would be dealt with as treason[70] or by way of the Riot Act 1714 (1 Geo 1 St 2 c 5), which was enacted in response to various internal and external threats to the fledgling Hanoverian regime, such as the Jacobites.[71] The Riot Act created felony offences of failing to disperse an hour after a proclamation had been read and of demolishing, or beginning to demolish, certain buildings, including churches and dwellings.[72] In addition, the reading of the Riot Act enabled the application of 'a kind of modified martial law', under which rioters were transformed not just into felons but also traitors against the Crown who therefore could be lawfully executed.[73] Even after the enactment of the Riot Act, however, prosecutors in the eighteenth century still favoured the charge of riot at common law, with the felony offence being reserved for the most serious cases.[74]

Although the Riot Act has now been repealed,[75] and common law riot replaced with a statutory offence in the 1986 Act, the offences of treason and treason felony remain on the statute books.[76] Incitement of violent conduct to overthrow the state can also constitute sedition.[77] If attempts to violently overthrow the state should be criminalised, it remains the case that treason, treason felony or sedition, not riot, most appropriately convey censure for that act.

It is debatable, of course, whether the offences of treason and sedition are necessary today. The Law Commission in 1977 thought

---

69 Silver (n 68 above) 15–17; Randall (n 63 above) 17, 20–23, 42–43; Steve Hall and Simon Winlow, 'The English riots of 2011: misreading the signs on the road to the society of enemies' in David Pritchard and Francis Pakes (eds), *Riot, Unrest and Protest on the Global Stage* (Palgrave Macmillan 2014); R Quinault and J Stevenson (eds), *Popular Protest and Public Order: Six Studies in British History 1790–1920* (George Allen & Unwin 1974) 26.

70 Smith (n 7 above) 1–2; Williams (n 66 above) 248–250; John Baker, *The Oxford History of the Laws of England: Volume VI 1483-1558* (Oxford University Press 2003) 584–585.

71 Griffin (n 66) 169; Randall (n 63 above) 2–3, 24–25; Richard Vogler, *Reading the Riot Act: The Magistracy, the Police and the Army in Civil Disorder* (Open University Press 1991) 1.

72 Riot Act 1714, ss I–II, IV.

73 Ibid ss I–III.

74 W Nippel, '"Reading the Riot Act": the discourse of law enforcement in 18th century England' (1985) 1 History and Anthropology 401, 415.

75 Statute Law (Repeals) Act 1967 s 1, sch 1 pt 5.

76 Treason Act 1351; Treason Felony Act 1848. See Supperstone (n 64 above) 230–234; Law Commission, *Codification of the Criminal Law: Treason, Sedition and Allied Offences* (Law Com Working Paper No 72 1977) paras 39, 57.

77 Supperstone (n 64 above) 234–240.

that, although sedition was no longer necessary,[78] a specific offence to penalise conduct aimed at overthrowing the constitutional government was needed to reflect the nature of the act as a crime against the state.[79] Considering the issue again in 2008, it questioned the need for treason offences in peacetime and referred to the use of public order offences to deal with 'serious civil unrest', but noted that treason laws could be 'simplified and pruned'.[80] The implication is that some form of treason law is still necessary.

As for the specific issue of riots threatening the state, there has not been a single instance of a riot or other public assembly in England and Wales in the post-war era that has seriously threatened to overthrow the state,[81] even indirectly, for example by causing a collapse in the rule of law. This is not to say that there are no rioters or protestors who wish to overthrow the current system of government, such as some anarchists or communists, but the aims of these people tend to be quite tangential to what actually fuels most contemporary rioting. This is evident from an examination of the largest protests in England and Wales's contemporary history, such as the march against the Iraq War in 2003 and the tuition fee protests of 2011, and of the most serious riots, such as the poll tax riots in 1990, the Bristol riot of 1980, the Brixton riot of 1981, the Bradford riot of 2001 and the England riots of 2011, not one of which was aimed at revolution or extra-constitutional governmental change.

The recent Capitol Hill 'riot' in the US cautions against concluding that there will never be a riot that threatens to overthrow even an established democratic state. However, the mischief in those circumstances is not the riot as such, but the attempted coup. Riot is not the right label for attempting to violently overthrow the state. The elements of the riot offence almost pale into insignificance compared to the gravity of the treason offences, 'the most serious of all criminal offences':[82] to establish riot, a group of only 12 rioters is required, of which only the offender needs actually to use violence, and no harm or fear needs to have been caused. Riot's maximum sentence of 10 years is also trivial compared to the maximum of life imprisonment for treason by levying war and treason felony.[83]

78 Law Commission, *Codification of the Criminal Law: Treason, Sedition and Allied Offences* (n 76 above) para 77.

79 Ibid 59, 61.

80 Law Commission, *Tenth Programme of Law Reform* (Law Com No 311 2008) paras 2.28–2.30.

81 See eg Townshend (n 6 above) ch 7.

82 Law Commission, *Codification of the Criminal Law: Treason, Sedition and Allied Offences* (n 76 above) para 21.

83 Supperstone (n 64 above) 233.

Turning to the 'inchoate' interest of preventing public fear of harm, the riot offence is inchoate in that it proscribes conduct (the use of unlawful violence as part of a group of 12 or more persons) that is likely to result in a particular outcome (public fear of physical harm). As is generally the case with inchoate offences,[84] it does not matter whether the outcome occurs. The violence needed for riot does not necessarily need to harm a person or property (nor does it need to have been intended to do so); it merely needs to have been capable of causing such harm.

It is noteworthy that the offence is purportedly aimed at the *fear* of physical harm, not at the physical harm itself or even the *risk* of the physical harm, unlike offences of risk creation such as drunk driving or criminal damage endangering the life of another. As Card puts it, the need for the offences of riot, violent disorder and affray seems to be based on so-called 'group offending' 'caus[ing] particular fear in ordinary members of the public and increased difficulties for the police'.[85] The gravamen of these three offences, and what marks them out as public order offences, is said to be their 'capacity to put in fear a notional bystander of reasonable firmness'.[86] The Law Commission expressly refers to 'terror' marking the character of common law affray as an offence against public order.[87]

But is it really the case that the public fear caused by a large group of people threatening or using violence is what marks riot as distinct from offences against the person or property? For one thing, since the 'person of reasonable firmness present at the scene' is hypothetical,[88] no one actually needs to be put in fear of harm. For example, an armed burglary of an unoccupied rural cottage carried out by 12 persons who violently broke down the door could constitute riot, even if no one witnessed the burglary.[89] Similarly, participants in a gang fight within a private dwelling could be convicted of riot even if they were all willing participants and there were no bystanders.[90]

It seems then that, although the commission of the riot offence will normally involve situations where there is public fear and alarm, the offence is not, strictly speaking, directed at protecting against fear as such. As Ashworth puts it, the 1986 Act grants an 'express

84 Ashworth (n 41 above) 9.

85 Card (n 7 above) 83.

86 Ibid 93.

87 Law Commission, *Offences Relating to Public Order* (n 10 above) para 3.29. 'Terror' or 'alarm' was required for riot at common law: Supperstone (n 64 above) 131; Thornton (n 9 above) 12–13.

88 1986 Act, s 1(4).

89 Cf *London and Lancashire Fire Insurance Company, Ltd v Bolands, Ltd* [1924] AC 836 (HL).

90 Adapting an example from Smith (n 7 above) 78.

dispensation' that any member of the public be put in fear, which he argues 'virtually undermines' the rationale for the riot offence.[91] A simpler explanation is that preventing fear is not the real rationale. Notably, the Law Commission, when rejecting the notion that it should be necessary for an actual bystander to be put in fear, stated that 'the function of the bystander is really to act as a measure of the requisite degree of violence'.[92]

Moreover, there are 'mainstream' offences which also purport to address fear, such as common assault (an act which causes another to apprehend immediate unlawful violence), public nuisance (as discussed in the previous section), threatening to kill,[93] threatening to damage or destroy property[94] and threatening violence to secure entry into occupied premises.[95] The last three, like riot, do not require anyone to be put in fear.[96] The existence of these offences further reduces any distinctiveness of the riot offence to the sole fact that it involves a group of at least 12 people.

The real mischief targeted by the riot offence is not public fear of harm, nor attempts to overthrow the state, but group violence, with the emphasis being on the 'group' element. Textbooks on public order law almost invariably quote a passage from Sachs LJ's judgment in *Caird* which is said to identify the real crux of the riot offence;[97] indeed, it is quoted four times in the Law Commission's final report:[98]

> [Riot] derives its great gravity from the simple fact that the persons concerned were acting in numbers and using those numbers to achieve their purpose ... The law of this country has always leant heavily against those who, to attain such a purpose, use the threat that lies in the power of numbers.[99]

This theme was later picked up by Lord Lane in a case relating to the 1984 miners' strike:

> It must have been obvious to all those participating in the picketing that their presence in large numbers was part of the intimidation and threat. It must have been clear to them that their presence would, at the least, encourage others to threats and/or violence, even if they themselves said nothing.

---

91 Ashworth (n 41 above) 17.
92 Law Commission, *Offences Relating to Public Order* (n 10 above) para 3.32.
93 Offences Against The Person Act 1861, s 16.
94 Criminal Damage Act 1971, s 2.
95 Criminal Law Act 1977, s 6.
96 Ashworth (n 41 above) 9.
97 Smith (n 7 above) 76; Thornton (n 9 above) 7.
98 Law Commission, *Offences Relating to Public Order* (n 10 above) paras 2.10, 6.4, 6.7, 6.11.
99 *R v Caird* (1970) 54 Cr App R 499 (CA), 505–507.

> One of the first requirements of any civilised society is that bullying should not succeed, that mere physical strength or strength of numbers should not be permitted to coerce the weaker or the fewer in number.[100]

As Blake argues, this is equivalent to suggesting that gathering in large numbers amounts to a criminal offence.[101] Although there is nothing inherently dangerous or harmful about a crowd,[102] the broad definition of 'violence' for the purposes of riot, discussed in the previous section, means that the slightest disturbance by a group of at least 12 people – pushing at police cordons; shouting in an intimidating fashion; throwing drink cans and plastic bottles[103] – can theoretically lead to a riot charge. The emphasis of the offence is very much on the weight of numbers, not the violent conduct. In fact, of all the charges brought in relation to the 1984 miners' strike, only 8.4 per cent were for crimes of *actual* violence, namely assaulting a police officer, actual bodily harm, grievous bodily harm, murder, wounding and possession of an offensive weapon.[104]

Somewhat incongruously, Lord Lane continued in his judgment as follows, seeming to imply that riot is but assault writ large:

> This requirement is exemplified inter alia by the common law offence of assault. An assault is any act by which the defendant intentionally, or recklessly, causes the victim to apprehend immediate unlawful violence. There is no need for it to proceed to physical contact.[105]

THE USE AND ABUSE OF THE RIOT OFFENCE

I have concluded that the riot offence is neither instrumentally necessary for adequate punishment nor justified on the grounds of protection

---

100 *R v Mansfield Justices, ex parte Sharkey* [1985] 1 QB 613 (QB), 627.

101 Nick Blake, 'Picketing, justice and the law' in Bob Fine and Robert Millar (eds), *Policing the Miners' Strike* (Lawrence & Wishart 1985) 110; Nadine El-Enany, '"Innocence charged with guilt": the criminalisation of protest from Peterloo to Millbank' in David Pritchard and Francis Pakes (eds), *Riot, Unrest and Protest on the Global Stage* (Palgrave Macmillan 2014) 75.

102 George Gaskell and Robert Benewick, 'The crowd in context' in George Gaskell and Robert Benewick (eds), *The Crowd in Contemporary Britain* (Sage 1987) 12–13; John Edwards, Robin Oakley and Sean Carey, 'Street life, ethnicity and social policy' in George Gaskell and Robert Benewick (eds), *The Crowd in Contemporary Britain* (Sage 1987) 76–77; Clifford Stott et al, 'Tackling football hooliganism: a quantitative study of public order, policing and crowd psychology' (2008) 14 Psychology, Public Policy and Law 115, 136.

103 As occurred at the beginning of the poll tax riots: Ian Hernon, *Riot! Civil Insurrection from Peterloo to the Present Day* (Pluto Press 2006) 239–240.

104 Janie Percy-Smith and Paddy Hillyard, 'Miners in the arms of the law: a statistical analysis' (1985) 12 Journal of Law and Society 345, 350.

105 *R v Mansfield Justices, ex parte Sharkey* (n 100 above) 627.

against public fear and overthrow of the state. Instead, the offence, at its core, targets the supposed mischief of gathering in a group of 12 or more and exhibiting behaviour deemed to be 'violent'. In this section, I examine how the vagueness of the definition of the offence enables its abuse and serves less obvious, ideological functions.

An implication of the riot offence's lack of a need for an actual bystander to be put in fear is that riot charges and convictions do not depend on witness testimony or victim statements and are highly dependent on police accounts of events. As such, all other things being equal, the police are incentivised to prefer public order charges, even if charges of criminal damage or assault are justified, because of the lower evidential burden.[106] As Ashworth argues, there is 'little doubt that the public order … offences have been defined so as to favour the convenience of prosecutors' (and, by extension, the police).[107]

This factor, along with the low threshold for 'violence', is common to all three group disorder offences (riot, violent disorder and affray). These offences are therefore highly susceptible to deliberate abuse at worst, such as to curb political dissent, and careless misjudgement at best. Misjudgement by the police has certainly played a significant role in many riots in England, fuelled by an obstinate refusal to accommodate any version of public 'order' other than its own and exacerbated by the much-vaunted operational independence and increasing 'professionalisation' of the police.[108] As a police spokesman reportedly said during the Brixton riot in 1981: 'The police will not withdraw. The only people who control the streets of London are the Met.'[109] The same mindset is evident time and time again: the picketing miners at Orgreave were said to have 'no right to be there'; during the Toxteth riot of 1981, a warning was issued for 'law-abiding people [to] keep off the streets'.[110] More recently, the Commissioner of the Metropolitan Police defended the heavy-handed policing of a vigil on Clapham Common for a murdered woman, Sarah Everard, as follows: 'I don't think anybody who was not in the operation can actually pass a detailed comment on the rightness and wrongness.'[111]

However, the police view on 'rightness and wrongness' is not the only version of public order. The notions of 'order' and 'disorder' are

106 Andrew Sanders, Richard Young and Mandy Burton, *Criminal Justice* 4th edn (Oxford University Press 2010) para 3.4.2.

107 Ashworth (n 41 above) 17. See also Sanders et al (n 106 above) para 3.4.2.

108 Townshend (n 6 above) 15, 139, 191–202.

109 Paul Gordon, '"If they come in the morning …" The police, the miners and black people' in Bob Fine and Robert Millar (eds), *Policing the Miners' Strike* (Lawrence & Wishart 1985) 164.

110 Scraton (n 9 above) 390.

111 'Sarah Everard: Met Police chief will not resign over vigil scenes' (*BBC News* 14 March 2021).

highly contested and 'historically and politically contingent', being 'intimately connected with the dominant power relationships of a society'.[112] Minority subcultures, such as those of the communities in Brixton in 1981 or Broadwater Farm in 1985 (or even football fans),[113] have their own norms and form of order. When these norms clash with the norms of the dominant culture, physical conflict may result.[114] Furthermore, rioting is not inherent to collective behaviour but develops when a form of order seen as illegitimate is imposed, through police coercion, on a crowd.[115] For example, the Broadwater Farm riot of 1985 was triggered by a disproportionate police response, including the premature deployment of riot gear, in response to a march in protest at the death of a local woman during a police raid.[116] Such marches were a 'conventional mode of peaceful protest in that community' and would normally follow a script well-known to both residents and the police. In this instance, however, it was perceived and portrayed by the police as a 'menacing incident'.[117] Notably, it was another 'ritual' march through Broadwater Farm to the police station in protest of the shooting of Mark Duggan and a heavy-handed attempt by police to disperse the resulting gathering that triggered the 2011 English riots.[118]

More significantly, the nebulousness of the group disorder offences can enable their abuse for the purpose of legitimising the crushing of political dissent. For example, the striking miners of 1984 were depicted as 'violent mobs' and 'invading hordes', despite the evidence of their violence being thin, as discussed in the previous section.[119] Rowdy

112 Chris Cunneen and Mark Findlay, 'The functions of criminal law in riot control' (1986) 19 Australian and New Zealand Journal of Criminology 163, 166. See also Wells and Quick (n 41 above) 228; Townshend (n 6 above) 15.

113 Football disorder tends to occur when coercive police intervention leads to 'ordinary' fans uniting with 'hooligans' around a perception of victimisation: Stott et al (n 102 above).

114 Philip Norton, 'Introduction' in Philip Norton (ed), *Law and Order and British Politics* (Gower 1984) 5.

115 Clifford Stott and John Drury, 'Contemporary understanding of riots: classical crowd psychology, ideology and the social identity approach' (2017) 26 Public Understanding of Science 2, 12; Stott et al (n 102 above) 136. See also Benjamin Bowling, Robert Reiner and James Sheptycki, *The Politics of the Police* 5th edn (Oxford University Press 2019) 89–90, 93–94.

116 Clive Bloom, *Violent London: 2000 Years of Riots, Rebels and Revolts* (Sidgwick & Jackson 2003) 443.

117 Wells and Quick (n 41 above) 229; Matthew Moran and David Waddington, *Riots: An International Comparison* (Macmillan 2016) 132–133.

118 Moran and Waddington (n 117 above) 116, 131–134.

119 Bob Fine and Robert Millar, 'Introduction: The Law of the Market and the Rule of Law' in Bob Fine and Robert Millar (eds), *Policing the Miners' Strike* (Lawrence & Wishart 1985) 18–19; Blake (n 101 above) 109; Percy-Smith and Hillyard (n 104) 350.

behaviour, such as 'pushing and shoving on the picket line',[120] was equated with violence. The strikes culminated in the well-documented 'Battle of Orgreave', a 'series of set piece battles' between the police and the miners.[121] Notoriously, the subsequent trial of 15 miners for riot collapsed after police evidence was discovered to have been fabricated.[122]

By contrast, in the previous year, the same police force had allowed to pass uneventfully the 'noisy and extremely boisterous' 'Thatcher Unwelcoming' demonstration led by local politicians and local leaders in Sheffield in 1983, despite there being some 'minor hostility' in the form of throwing of foodstuffs.[123] One factor for the difference in treatment appears to be 'the perceived legitimacy of the demonstration from a senior police perspective'.[124]

As Wells and Quick argue, all this demonstrates

> the malleability of the notion of disorder as a threat to state authority and the ways in which it can be appealed to reinforce punitive state reactions to forms of behaviour which, taken as individual instances, would not be seen in nearly such threatening terms. In such contexts, we can see that the use of the term 'public' signifies not a particular sphere of activity (already hard to define) but rather the conception of order which prevails: that of the state or particular powerful groups within it.[125]

Public order offences serve symbolically to affirm the authority of the state and the 'agencies of control',[126] as well as to legitimise the state's, or the police's, conception of order. If the police are merely enforcing the law, the 'rightness' of their coercive actions cannot be questioned.

Furthermore, the existence and use of public order offences depoliticises situations of 'disorder', allowing the state to portray riot participants as mere criminals and to ignore any underlying

120 Fine and Millar (n 119 above) 18–19.

121 Cathie Lloyd, 'A national riot police: Britain's "third force"?' in Bob Fine and Robert Millar (eds), *Policing the Miners' Strike* (Lawrence & Wishart 1985) 66; David Waddington, 'Policing political protest: lessons of best practice from a major English city' in David Pritchard and Francis Pakes (eds), *Riot, Unrest and Protest on the Global Stage* (Palgrave Macmillan 2014) 18; John Alderson, *Principled Policing: Protecting the Public with Integrity* (Waterside Press 1998) 154–158.

122 Joanna Gilmore, 'Lessons from Orgreave: police power and the criminalization of protest' (2019) 46 Journal of Law and Society 612.

123 David Waddington, Karen Jones and Chas Critcher, 'Flashpoints of public disorder' in George Gaskell and Robert Benewick (eds), *The Crowd in Contemporary Britain* (Sage 1987) 165.

124 Waddington (n 121 above) 14; Waddington et al (n 123 above) 166–170.

125 Wells and Quick (n 41 above) 226.

126 Ibid 211.

grievances.[127] Focus shifts from the grievances of the riot participants to the punishment that they deserve.[128] The criminal law, for the purposes of establishing guilt, does not look at motive and, as such, it is said that 'we have no special law for protestors'.[129] However, this 'obscures the socio-political reality of [public disorder] and distracts public attention from the broad base of such public behaviour'.[130] The behaviour constituting the riot offence is decontextualised from its social context,[131] even though the social context is morally relevant.[132] It also allows the state to tackle political dissent without appearing to do so.[133] The case of the striking miners in 1984 is a case in point: their motives were obviously political and the then Prime Minister Margaret Thatcher herself described them as 'an organised revolutionary minority'.[134] Yet she also depicted them as mere criminals.[135] More remarkably, she expressed the same attitude towards disorder in Northern Ireland: 'a crime is a crime is a crime'.[136]

The 2011 English riots provide a potential complication to this argument. As discussed earlier, riot formed a very small proportion of the crimes recorded and the convictions. More generally, only a fifth of the defendants proceeded against at magistrates' courts for all offences relating to the 2011 riot were convicted for 'violent disorder' offences (defined broadly to include public order offences as well as other offences such as common assault and assaulting a constable).[137] This is even though some of the acquisitive offences charged would likely also have fallen under riot or violent disorder. Indeed, guidance issued to prosecutors by the Crown Prosecution Service stated that 'the offence of riot merits serious consideration'.[138]

However, the low prevalence of convictions for 'violent disorder' may be due to the 2011 riots being widely perceived as criminal. Although sparked by poor police communication following a police shooting, the riots were quickly associated predominantly with looting and have

127 Ibid 229.
128 Gilmore (n 122 above) 638.
129 Smith (n 7 above) 4.
130 Cunneen and Findlay (n 112 above) 170.
131 El-Enany (n 101 above) 89.
132 Wells and Quick (n 41 above) 214.
133 El-Enany (n 101 above) 79, 85.
134 Fine and Millar (n 119 above) 2.
135 Ibid.
136 Paddy Hillyard, 'Lessons from Ireland' in Bob Fine and Robert Millar (eds), *Policing the Miners' Strike* (Lawrence & Wishart 1985) 178.
137 Ministry of Justice (n 31 above) Table 1.2.
138 Lightowlers and Quirk (n 24) 71.

subsequently been described as 'consumerist' riots.[139] As the then Prime Minister David Cameron said: 'This was not a political protest or a riot about politics. It was common or garden thieving, robbing and looting.'[140] Even scholars of rioting remain divided on whether the riots were 'political'.[141] Absent any popular association of the riots with political grievances, there was no need for police and prosecutors to choose the relatively riskier charges of riot and violent disorder. By contrast, most of the defendants charged following the 2001 Bradford riots were convicted of riot.[142]

Riot, as an indictable offence, is always tried before a jury, which may serve as a last line of defence against politically motivated riot charges. Although only tentative conclusions can be drawn from the data, given the small number of riot trials in each year, the figures on acquittals seem to suggest that the choice to charge riot is a risky one for prosecutors. Out of the 78 defendants tried for riot between 2004 and 2013 (inclusive),[143] 21 (27%) were acquitted. This rate is about the same as the acquittal rates at the Crown Court in the same period for violent disorder (26%), but higher than those for public order offences (not limited to those in the 1986 Act) (18%) and all offences (20%).[144]

Although the statistics must be read with caution given the small sample size, they align with the notorious difficulty of securing convictions for riot at common law.[145] I have already mentioned the collapse of the riot charges against the miners at the 'Battle of Orgreave'. There were less well-known failures to convict other striking miners, including 13 riot acquittals by a Sheffield jury and eight acquittals in Nottingham for riotous assembly and affray.[146] The collapse of the riot trial relating to the Bristol riot of 1980 is also well-known: of the 12 defendants tried, eight were acquitted (including three after a direction from the judge). The trial collapsed as the jury were deadlocked on the remaining defendants, and a retrial was not sought.[147] Given that jury

139 Tim Newburn, 'The 2011 England Riots in Recent Historical Perspective' (2015) 55 British Journal of Criminology 39, 53–56.

140 Quoted in ibid 51.

141 Ibid; Sadiya Akram, 'Recognizing the 2011 United Kingdom riots as political protest: a theoretical framework based on agency, habitus and the preconscious' (2014) 54 British Journal of Criminololgy 375.

142 Lightowlers and Quirk (n 24) 71.

143 See Table 2. The corresponding data for 1987 to 2003, which are not available online, could not be obtained before finalisation of this article. As stated above, there were no riot trials between 2014 and 2020 (inclusive).

144 Ministry of Justice (21 May 2015) (n 27 above).

145 Thornton (n 9 above) 16–17; Williams (n 66 above) 240.

146 Hernon (n 103 above) 235; Thornton (n 19 above) 26–27.

147 Martin Kettle and Lucy Hodges, *Uprising! The Police, the People and the Riots in Britain's Cities* (Pan Books 1982) 34–38; Thornton (n 19 above) 26.

deliberations are secret, we may never know definitively why juries acquit riot defendants. The technical requirements may be one factor, with the requirement to prove a common purpose being described by Lord Scarman as a matter of 'great forensic confusion'.[148] It may not be too much of a stretch, however, to suggest that, in politically charged circumstances, modern juries are continuing a long historical tradition of jury sympathy for rioters.[149] Three of the Bristol jurors even joined in the post-trial celebrations![150]

Finally, English criminal law, with its emphasis on individual responsibility, is ill-suited to deal with riot, an essentially collective activity.[151] As Cunneen and Findlay argue, there is a contradiction in relying on the collective nature of the behaviour in justifying the offence but denying its relevance when determining individual responsibility.[152] This is compounded by the fact that only rioters who are arrested can be convicted. This issue does affect all crime – most offenders are not caught and therefore not charged and convicted – but the unfairness is particularly acute in the context of group disorder offences, given that the very gravity of the offences is derived from the presence of a group.

In *Caird*, Sachs LJ disposes of what he calls the 'Why pick on me?' argument as follows:

> [O]n these confused and tumultuous occasions each individual who takes an active part by deed or encouragement is guilty of a really grave offence by being one of the number engaged in a crime against the peace. It is, moreover, impracticable for a small number of police when sought to be overwhelmed by a crowd to make a large number of arrests ... Those who choose to take part in such unlawful occasions must do so at their peril.[153]

As with the more famous passage from his judgment, the emphasis is on the group element: rioters deserve punishment not because of violence, but because as a collective they have breached 'the peace' and 'overwhelmed' the police. Sachs LJ rejects the defendants' contention that their acts should be regarded in isolation, but goes on to consider the appropriate sentence for each individual defendant.[154]

---

148 Quoted in Thornton (n 9 above) 12.

149 See eg Nippel (n 74 above) 417; Randall (n 63 above) 26–28; Bowling et al (n 115 above) 67.

150 Kettle and Hodges (n 147 above) 38.

151 Ralf Dahrendorf, *Law and Order* (Stevens & Sons 1985) 33; Smith (n 7 above) 2–3.

152 Cunneen and Findlay (n 112 above) 165. See also Wells and Quick (n 41 above) 212.

153 *R v Caird* (n 99 above) 506–507.

154 Ibid 507–509.

On a related note, it is interesting that acting under 'mass suggestion' was a mitigating factor under the Italian and Cuban penal codes,[155] perhaps reflecting the influential but now-discredited theory of Le Bon that posited the crowd as having a suggestible and primitive mind of its own that subsumed individual conscious personalities and individual rationality.[156] Hints of this mindset were discernible in reactions to the 2011 English riots: David Lammy, MP for Tottenham, where the riots first broke out, referred to the rioters as 'mindless', whilst Met Commander Adrian Hanstock referred to 'mindless thugs'.[157] The portrayal of crowds as 'mindless' serves to legitimise state repression and to delegitimise political grievances – civilisation must be protected from the pathology of mindlessness.[158] English criminal law, however, adopts the opposite extreme of pretending that individuals have perfect free will.[159] The crowd's influence is no excuse and, like intoxication, is even an aggravating factor.[160]

CONCLUSION

In a 1991 polemic, P A J Waddington takes issue with the 'critical consensus' of academia that is critical of the police's role in enforcing public order:

> In the event of widespread racist violence against ethnic minorities, there is no doubt that those who now complain about the policing of public order would be anxious to see the police take effective and, if necessary, forceful action, because now the police are not playing the part of oppressive ogres but are the equivalent of the 7th Cavalry.[161]

Waddington's point is that police tactics are 'a *means*, not an end' that can be used both to 'stifle legitimate protest' and 'protect vulnerable minorities'.[162] The same argument could apply to public order law. It is similar to the sentiment encapsulated in the adage that 'we have no special law for protestors'. The sentiment can also be found in the argument that, unlike in the eighteenth century, when riots were a

155 Thornton (n 19 above) 23–24; Hermann Mannheim, *Comparative Criminology: A Text Book* (Routledge & Kegan Paul 1965) 655.

156 Tim Newburn, 'The causes and consequences of urban riot and unrest' (2021) 4 Annual Review of Criminology 53, 55–56; Stott and Drury (n 115 above) 9.

157 Stott and Drury (n 115 above) 3.

158 Ibid 9.

159 Wells and Quick (n 41 above) 213–214.

160 Sentencing Council, 'Aggravating and mitigating factors'.

161 P A J Waddington, *The Strong Arm of the Law: Armed and Public Order Policing* (Clarendon Press 1991) 251.

162 Ibid.

form of 'articulate' protest that the elite listened to,[163] Britain now has democratic processes by which people can choose their representatives as well as the right to protest peacefully. The implication is that crowd violence no longer has any place in our political system, whatever the grievances.[164] 'A crime is a crime is a crime.'

In this article, I have not considered directly whether riots can be legitimate forms of protest; I will leave questions of the morality of or normative justifications for rioting to the political theorists and ethicists.[165] Waddington is also right to point out that not all riots are driven by a desire for progressive social change. In fact, some are just spontaneous outbreaks of violence not driven by any social or political grievance.[166] Furthermore, riots are hugely damaging, costly and traumatic events, and I have sought not to minimise these effects.

Nevertheless, there are significant problems with the offence of riot. First, it is unnecessary given the range of other offences available to punish riot participants, some of which have equal or higher maximum penalties. Second, the mischief that riot tackles is neither protection of the public from fear nor protection against overthrow of the state. The real 'mischief' is the gathering of people in groups, but there is nothing inherently dangerous or harmful about a crowd. Indeed, the right to freedom of assembly is fundamental to a healthy democracy. The breadth of the 'violence' needed to constitute riot allows the law to be enforced at the slightest hint of a disturbance. Third, because of this and the lack of the need for anyone to be actually harmed, the offence is ripe for abuse. If 'we have no special law for protestors', then why do we have a riot offence that is by its nature a discretionary law predicated on a particular notion of 'order' and 'disorder', that of those with power, such as the state and the police?

Given all this, the riot offence should be abolished (although the definition could be retained for the purposes of the Riot Compensation Act 2016). This is not the same as saying that riot participants should not be punished if they harm persons or damage property. That is what the offences against the person or property are for. Arrests, charges and convictions should not be based mainly on individuals being in a group, but on each individual's actions and the harm caused. The room for political value judgements in arrest and charging decisions would shrink: football hooligans and socially aggrieved rioters would

163 Silver (n 68 above) 17–19, 23.

164 Kettle and Hodges (n 147 above) 17; Stuart Hall, *Drifting into a Law and Order Society* (The Cobden Trust 1980) 9.

165 See eg Avia Pasternak, 'Political rioting: a moral assessment' (2018) 46 Philosophy and Public Affairs 384; Jonathan Havercroft, 'The British Academy Brian Barry Prize Essay: why is there no just riot theory?' (2021) 51 British Journal of Political Science 909.

166 Newburn (n 156 above) 63.

be punished equally if they commit one of the 'mainstream' offences in the course of a riot. Charging the 'mainstream' offences would also enable a more nuanced discussion of riots: actual violence to persons and property would not be condoned, but any grievances underlying riots would not be obscured by the criminal label of 'riot'. It is true that some acts that would constitute 'violence' under the riot offence may fall through the gaps if there is no riot offence. I discussed this issue and ventured that these acts would be few in number and queried whether they would actually be 'violent'. Why should 'violence' that falls short of even common assault be penalised, and penalised so heavily, other than because of the imagined dangerousness of the crowd?

Riots happen – frequently.[167] As we look back more than a decade to the 2011 English riots, we undoubtedly hope that the destruction wrought is not repeated. But when the next riot does happen, we can do better than to repeat the mistake of wielding the offence of riot.

167 See eg Bloom (n 116 above); Hernon (n 103 above); Vogler (n 71 above).

Northern Ireland
Legal Quarterly
Winter Vol. 73 No. 4 (2022) 742–767
Article DOI: 10.53386/nilq.v73i4.525

Coercive control, legislative reform and the Istanbul Convention: Ireland's Domestic Violence Act 2018

Judit Villena Rodó

University of Galway

Correspondence email: j.villenarodo1@universityofgalway.ie

ABSTRACT

Coercive control is a concept increasingly being used in legal and policy responses to intimate partner violence. This article examines this concept in light of Ireland's obligations under the Council of Europe Convention on Preventing and Combating Violence Against Women and Domestic Violence (Istanbul Convention) to prevent and combat domestic violence, including psychological violence (arts 3 and 33). First, it analyses the interpretation of article 33 by the Council of Europe Group of Experts on Action against Violence Against Women and Domestic Violence (GREVIO) in country monitoring reports. Second, it examines Ireland's coercive control offence, comparing it to legislative developments in the United Kingdom (UK). Third, it examines potential theoretical and practical concerns arising from the application of the offence, drawing from literature on the criminalisation of coercive control in the UK. It argues that concerns regarding the practical application of the offence may be relevant to Ireland.

Keywords: coercive control; Istanbul Convention; violence against women; human rights law.

INTRODUCTION

Ireland's newest domestic violence legislation, the Domestic Violence Act 2018 (DVA 2018) strengthens the legal framework relating to domestic violence in Ireland through, *inter alia*, the introduction of additional civil law protection measures and the criminalisation of coercive control and forced marriage.[1] The introduction of the offence of coercive control in Ireland follows similar developments elsewhere. Since 2015, new offences of coercive control have been introduced in England & Wales, Scotland and Northern Ireland, as part of legislative and policy reforms to address domestic

1 Women's Aid and Monica Mazzone, *Unheard and Uncounted: Women, Domestic Abuse and the Irish Criminal Justice System* (Women's Aid 2019) 13; Domestic Violence Act 2018, ss 38, 39 (Ireland).

violence.[2] Coercive control refers to an often invisible form of harm, prevalent in patterns of domestic violence, that seeks to deprive the victim/survivor of her liberty and personhood through tactics such as intimidation, threats, isolation or control.[3]

This article analyses the enactment of the coercive control offence in Ireland and interrogates the substantive achievements, shortcomings and opportunities posed by the DVA 2018's offence in light of the Council of Europe Convention on Preventing and Combating Violence Against Women and Domestic Violence (Istanbul Convention), ratified by Ireland in 2019.[4] It argues that, although the introduction of this new offence is a significant step towards justice for victims/survivors of coercive control, there is a need for further procedural guidance and training for the judiciary, An Garda Síochána (Ireland's police force) and for prosecutors to ensure its effective implementation and further cultural change. Thus, this article begins by briefly introducing the phenomenon of coercive control.[5] The piece then examines the standards set out in the Istanbul Convention concerning the criminalisation of psychological violence, and state parties' obligations relating to access to justice for victims/survivors of such abuse. The article proceeds to present an account of the issues arising regarding the introduction of the offence of coercive control in the parliamentary debates on the DVA 2018. To conclude, it reflects on the normative achievements, the missed opportunities of the DVA 2018 in relation to coercive control, and the impact that the implementation of this new offence may have on victims/survivors' access to effective remedies.

A BRIEF INTRODUCTION TO THE CONCEPT OF COERCIVE CONTROL

Contemporary understandings of domestic violence identify coercive control as its central element. Evan Stark and others developing this

2 Serious Crime Act 2015, s 76, as amended by Domestic Abuse Act 2021, s 68 (England & Wales); Domestic Abuse (Scotland) Act 2018, s 1; Domestic Abuse and Civil Proceedings Act (Northern Ireland) 2021, s 1. For an in-depth comparative analysis of the offences, see Vanessa Bettinson, 'A comparative evaluation of offences: criminalising abusive behaviour in England, Wales, Scotland, Ireland and Tasmania' in Marilyn McMahon and Paul McGorrery (eds), *Criminalising Coercive Control: Family Violence and the Criminal Law* (Springer 2019).

3 Evan Stark, *Coercive Control: How Men Entrap Women in Personal Life* (Oxford University Press 2007).

4 Council of Europe, Convention on Preventing and Combating Violence Against Women and Domestic Violence (1 August 2014). Ireland signed the Convention in 2015 and ratified it in 2019.

5 In line with the legislative reality in Ireland, this article discusses coercive control solely in the context of intimate partnerships.

argument have advocated for a turn away from a historical, social and legal conceptualisation of domestic violence as incidental, and assault and physical violence based.[6] Stark, whose work underpins most of the legislative reform in Europe to date, has described this form of abuse as a course of domination achieved through a combination of coercion tactics, including violence and intimidation, 'deployed to hurt and intimidate'; and control tactics, including isolation and attempts to regulate and confine victims/survivors.[7] Generally, state intervention on domestic violence continues to overlook this underlying and invisible form of harm, often referred to by victims/survivors as 'the worst part', creating an inadequate response and routinely failing victims/survivors accessing justice.[8]

Coercive control is a cyclical form of abuse, generally exerted over long periods of time, with a 'cumulative' effect on its victims/survivors.[9] It can cause severe physical and psychological harm, but primarily results in what Stark refers to as a 'hostage-like condition of entrapment'.[10] Thus, it directly impacts victims/survivors' dignity and personhood and, in turn, affects their 'autonomy, rights and liberties'.[11] A central aspect of the coercive control model, generally not reflected in the relevant legislation, is its gendered construction. Male perpetrators typically exploit gendered dynamics in designing the abuse, in which gender stereotypes are weaponised to harm the victim/survivor. Stark explains that this form of abuse commonly involves the regulation of conduct socially associated with femininity, such as care work (including cooking and cleaning).[12] He calls these 'patriarchal-like controls in personal life' and argues that, through the web of behaviours amounting to coercive

6 Stark (n 3 above); Paul McGorrery and Marilyn McMahon, 'Criminalising "the worst" part: operationalising the offence of coercive control in England and Wales' (2019) 11 Criminal Law Review 957; Evan Stark, 'Rethinking coercive control' (2009) 15 Violence Against Women 1509; Emma Williamson, 'Living in the world of the domestic violence perpetrator: negotiating the unreality of coercive control' (2010) 16 Violence Against Women 1412; Evan Stark, 'Looking beyond domestic violence: policing coercive control' (2012) 12 Journal of Police Crisis Negotiations 199.

7 Evan Stark, 'Re-presenting battered women: coercive control and the defense of liberty' (Violence Against Women: Complex Realities and New Issues in a Changing World, Montreal 2012) 7, 8.

8 Charlotte Bishop, 'Domestic violence: the limitations of a legal response' in Sarah Hilder and Vanessa Bettinson (eds), *Domestic Violence: Interdisciplinary Perspectives on Protection, Prevention and Intervention* (Palgrave Macmillan 2016) 66; McGorrery and McMahon (n 6 above).

9 Stark, 'Re-presenting battered women' (n 7 above) 7.

10 Ibid.

11 Ibid 5, 7.

12 Stark, *Coercive Control* (n 3 above) 211.

control, male perpetrators control 'women's enactment of everyday life'.[13] He further contends that the domination achieved over women at home reinforces their subordinate position in society, perpetuating patriarchy's constrainment of their agency.[14]

Typically, the primary method by which victims/survivors are targeted and subjected to coercive control is through the commission of individualised acts of abuse. However, given that coercive control takes place in the context of structural systems of power, such as sexism and racism, both identity and structural factors play a key role. Kristin Anderson, in theorising the gendered nature of coercive control, notes that it is important to consider victims/survivors' and perpetrators' 'individual characteristics' but also cultural and social structural aspects of gender inequality in understanding differences between cases of coercive control.[15]

THE ISTANBUL CONVENTION: PSYCHOLOGICAL VIOLENCE UNDER ARTICLE 33

The legal significance of the Istanbul Convention, as the first human rights treaty to comprehensively address both gender-based violence (GBV) and domestic violence, is widely recognised.[16] Indeed, Ireland's ratification of the Istanbul Convention served as one of the main drivers in improving the domestic legal framework against domestic violence. The Convention binds states parties to exercise due diligence regarding GBV.[17] In fulfilling this standard, states must take legal and other measures to 'prevent, investigate, punish and provide reparation' for acts of GBV covered in the Convention. This obligation applies to acts committed by non-state actors, including individuals.[18]

While developed and built upon in the Istanbul Convention, this is an obligation that is rooted in the established United Nations

13 Ibid 171, 172.

14 Ibid 172.

15 Kristin L Anderson, 'Gendering coercive control' (2009) 15 Violence Against Women 1444, 1447.

16 Ronagh J A McQuigg, *The Istanbul Convention, Domestic Violence and Human Rights* (Routledge Research in Human Rights Law 2019) 22. Other international and regional documents addressing GBV and domestic violence include: the African Union, Protocol to the African Charter on Human and Peoples' Rights on the Rights of Women in Africa (Maputo Protocol) (adopted 28 March 2003, entered into force 11 July 2003); and the Organisation of American States, Inter-American Convention on the Prevention, Punishment and Eradication of Violence Against Women (Convention Belem Do Pará) (adopted 9 June 1994, entered into force 5 March 1995).

17 Council of Europe (n 4 above) art 5.

18 Ibid art 5.

(UN) framework against violence against women, including the UN Convention on the Elimination of all Forms of Discrimination against Women (CEDAW 1979) and the UN Declaration on the Elimination of Violence Against Women (DEVAW 1993).[19] Even though the CEDAW does not expressly refer to violence against women, the Convention has been interpreted by the CEDAW Committee in its General Recommendations No 19 (1992) and No 35 (2017) as encompassing violence against women, in both the private and public domains, as a form of gender discrimination and, as such, a violation of human rights and fundamental freedoms.[20]

The Istanbul Convention takes a holistic approach and has been praised for providing a gendered, modern and comprehensive account of GBV, including domestic violence, as both a human rights violation in and of itself, as well as a component in wider discrimination against women.[21] In a similar fashion to the CEDAW and DEVAW definition of violence against women and domestic violence, the Istanbul Convention adopts a broad definition of domestic violence, under article 3, as 'all acts of physical, sexual, psychological or economic violence' occurring 'within the family or domestic unit or between former or current spouses or partners' regardless of whether the perpetrator and victim/survivor share a dwelling or not. Specifically, on psychological violence,[22] article 33 prescribes the obligation of state parties to take all 'necessary legislative or other measures' to ensure criminalisation of 'intentional conduct … seriously impairing a person's psychological integrity through coercion or threats'.[23] This express call in the Convention to criminalise psychological violence

19 United Nations, Convention on the Elimination of All Forms of Discrimination against Women (18 December 1979) vol 1249; United Nations, Declaration on the Elimination of Violence Against Women (20 December 1993) A/RES/48/104.

20 UN CEDAW Committee, General Recommendation No 19: Violence Against Women (1992), para 19; UN CEDAW Committee, General Recommendation No 35 on Gender-Based Violence Against Women, Updating General Recommendation No 19 (14 July 2017).

21 Dubravka Šimonović, 'Global and regional standards on violence against women: the evolution and synergy of the CEDAW and Istanbul Conventions' (2014) 36 Human Rights Quarterly 590, 604–603.

22 There is a disagreement between scholars regarding the value of equating coercive control to psychological violence. For instance, Stark conceptually disagrees with coercive control being approached as psychological violence, as doing so may risk understanding the phenomenon in light of 'mental processes' in detriment of its structural nature. See further Stark, *Coercive Control* (n 3 above) 11.

23 Council of Europe (n 4 above) art 33.

is unprecedented within the human rights framework of protection against domestic violence.[24]

Despite the prevalence of the language of coercive control in academic and civil society discourse at the time of Istanbul Convention's adoption,[25] references to this specific form of abuse are absent from the text of the Convention. However, it can be easily argued that the comprehensive definitions regarding domestic violence in the Convention confer due diligence obligations upon states to address coercive control, as analogous to psychological violence. The Convention's explanatory report supports this claim, as it submits that article 33 on psychological violence intends to capture an intentional course of conduct extending beyond discreet incidents or, in other words, 'an abusive pattern of behaviour occurring over time'.[26]

It is essential to mention that the explanatory report does not clarify what the reference in article 33 to 'seriously impairing' someone's psychological integrity entails, nor does it specify which behaviour, threats or coercion may amount to violence. To some extent, the Council of Europe Group of Experts on Action against Violence Against Women and Domestic Violence (known as GREVIO), which monitors state compliance with the Istanbul Convention, has clarified the scope of article 33 and states' obligations arising therein.[27] GREVIO has insisted on the importance of visibilising psychological violence, highlighting its connection to other forms of violence, such as physical or economic, and naming its severity as a violation of 'the victim's psycho-social integrity'.[28] GREVIO has explained that the drafters' intention behind article 33 was that 'any act causing psychological duress', which, notably, 'can take various forms such as isolation, excessive control and intimidation', should be punished.[29]

---

24 Paul McGorrery and Marilyn McMahon, 'Criminalising psychological violence in Europe: (non-)compliance with article 33 of the Istanbul Convention' (2021) 42 European Law Review 211, 215.

25 Stark, *Coercive Control* (n 3 above) and references therein.

26 Council of Europe, 'Explanatory Report to the Council of Europe Convention on Preventing and Combating Violence against Women and Domestic Violence' (11 May 2011) para 181.

27 Council of Europe 'About GREVIO – Group of Experts on Action against Violence Against Women and Domestic Violence' (nd). Under the Convention's art 66(2), its membership is composed of 10 to 15 members 'taking into account a gender and geographical balance, as well as multidisciplinary expertise'. The Group produces country monitoring reports, which evaluate the domestic measures taken to implement the Convention.

28 Council of Europe (n 26 above).

29 GREVIO, 'Baseline Evaluation Report on Legislative and Other Measures Giving Effect to the Provisions of the Council of Europe Convention on Preventing Combating Violence against Women and Domestic Violence (Istanbul Convention) Monaco' (27 September 2017) 113.

GREVIO's reports have clarified various issues regarding the behaviours falling within the scope of article 33. These can include minor instances of intimidation, including those at the early stages of the violence, not necessarily amounting to severe violence or threats.[30] Building on the idea of violence as a continuum, as acknowledged in the explanatory report, GREVIO has reinforced the notion that article 33 does not refer to separate and distinct assaults but rather a pattern extending in time and beyond single incidents.[31] Ongoing abuse, GREVIO has stated, is encompassed within article 33, even if the acts which constitute it do not 'necessarily reach the threshold of criminalisation'.[32] Referring to these behaviours, GREVIO used the terminology of coercive control in 2018[33] and has continued using it in monitoring states' compliance with article 33.[34] This interpretation is a clear confirmation that states have obligations which go beyond a historical and stereotypical understanding of domestic violence, including coercive control. GREVIO has reinforced this assertion by

30 GREVIO, 'Baseline Evaluation Report on Legislative and Other Measures Giving Effect to the Provisions of the Council of Europe Convention on Preventing and Combating Violence against Women and Domestic Violence (Istanbul Convention) Austria' (27 September 2017) para 144; GREVIO, 'Baseline Evaluation Report on Legislative and Other Measures Giving Effect to the Provisions of the Council of Europe Convention on Preventing and Combating Violence against Women and Domestic Violence (Istanbul Convention) Denmark' (24 November 2017) para 162.

31 GREVIO, 'Monaco' (n 29 above) para 113.

32 GREVIO, 'Baseline Evaluation Report on Legislative and Other Measures Giving Effect to the Provisions of the Council of Europe Convention on Preventing and Combating Violence against Women and Domestic Violence (Istanbul Convention) Turkey' (15 October 2018) para 215.

33 Ibid.

34 See, for instance, GREVIO, 'Baseline Evaluation Report on Legislative and Other Measures Giving Effect to the Provisions of the Council of Europe Convention on Preventing and Combating Violence against Women and Domestic Violence (Istanbul Convention) Finland' (2 September 2019) para 157; GREVIO, 'Baseline Evaluation Report on Legislative and Other Measures Giving Effect to the Provisions of the Council of Europe Convention on Preventing and Combating Violence against Women and Domestic Violence (Istanbul Convention) Belgium' (21 September 2020) para 152; GREVIO, 'Baseline Evaluation Report on Legislative and Other Measures Giving Effect to the Provisions of the Council of Europe Convention on Preventing and Combating Violence against Women and Domestic Violence (Istanbul Convention) Malta' (nd) para 164; GREVIO, 'Baseline Evaluation Report on Legislative and Other Measures Giving Effect to the Provisions of the Council of Europe Convention on Preventing and Combating Violence against Women and Domestic Violence (Istanbul Convention) Poland' (16 September 2021) para 199; GREVIO, 'Baseline Evaluation Report on Legislative and Other Measures Giving Effect to the Provisions of the Council of Europe Convention on Preventing and Combating Violence against Women and Domestic Violence (Istanbul Convention) Slovenia' (12 October 2021) para 245.

stating that offences which require the use of force or a serious threat of violence are inadequate to cover psychological violence under the Convention.[35] Similarly, GREVIO has referred to the importance of this obligation in aiding the social understanding and recognition of psychological violence as a crime and its consequential impact in increasing reporting of domestic abuse.[36]

Nonetheless, the obligation to criminalise psychological violence contained in the Convention is qualified, as article 33 is one of the six provisions into which states may enter reservations.[37] Article 78(2) of the Convention provides for the possibility of foregoing criminal sanctions against behaviours covered under article 33, if non-criminal sanctions are in place. Thus, where a state decides that criminalisation is not the appropriate route, it must provide effective remedies under civil law where it has failed to prevent the violence or protect the victim/survivor.[38]

Ireland's obligations under the Istanbul Convention complement and strengthen its existing commitments under the CEDAW Convention, including obligations in relation to GBV outlined in CEDAW's General Recommendations Nos 19 and 35. Both Conventions are complementary, and this synergy reinforces human rights standards relevant to violence against women.[39] Crucial for the implementation of all obligations linked to women's right to live free from violence, Ireland has an obligation to condemn all forms of discrimination and protect women's rights, especially those relating to their protection, in a non-discriminatory manner.[40]

IRELAND'S ENACTMENT OF THE DOMESTIC VIOLENCE ACT 2018

Ireland seems to have been particularly slow in providing legal remedies to victims/survivors of domestic violence. The enactment of new legislation protecting victims' rights, such as the Criminal Justice (Victims of Crime) Act 2017 and the DVA 2018, together with the ratification of the Istanbul Convention, appears to be a marked shift away from past shortcomings. This is a significant departure, albeit one rightly considered to be '100 years too late'.[41] As Louise Crowley has pointed out, previously to the DVA 2018, victim protection against

35 GREVIO, 'Austria' (n 30 above) para 144.
36 Ibid.
37 Council of Europe (n 4 above) art 78 (2) (3).
38 Ibid art 29.
39 Šimonović (n 21 above).
40 Council of Europe (n 4 above) art 4.
41 Seanad Deb 1 March 2017, vol 250, no 8.

domestic violence used to be contingent on the victim's willingness to seek 'civil remedies in the form of a barring order or safety order'.[42] This was due to the fact that, despite some elements of domestic violence being criminally sanctionable, for example, under the Non-Fatal Offences Against the Person Act 1997, state intervention did not prove to be 'sufficiently robust'.[43]

The DVA 2018 Bill, as initially introduced in the Seanad, Ireland's upper house, by former Minister for Justice and Equality Frances Fitzgerald, enjoyed cross-party support, but did not contain a provision on coercive control.[44] In the words of Minister Fitzgerald, the Bill aimed to 'consolidate and reform the law on domestic violence to provide better protection for victims'.[45] Reform was pressing for two main reasons. First, the domestic legal framework in place to address intimate partner violence lacked an effective and comprehensive protective framework for victims/survivors.[46] The inadequacy of the legal framework can be attributed, as Crowley suggests, to the historical and widespread reticence of the Irish state to intervene in family matters. In the family context, when it came to state intervention, property and privacy rights habitually operated against the protection of victims/survivors of domestic violence.[47] Second, the framework in place was insufficient to meet the standards set by the Istanbul Convention, including the obligations outlined in article 33. In theory, psychological violence and abuse could have been (inadequately) criminally punished under the harassment provision of the Non-Fatal Offences Against the Person Act 1997. As Crowley states, the operationalisation of the harassment provision would have taken place in a dire context where the legal minimisation of domestic violence was normalised, evidenced by a 'reluctance to charge offenders for criminal acts in the domestic context'.[48]

Despite the political will to legally consolidate and reform the state's domestic violence response, civil society submissions[49] and

42 Louise Crowley, 'Domestic violence law' in Lynsey Black and Peter Dunne (eds), *Law and Gender in Modern Ireland: Critique and Reform* (Bloomsbury 2019) 149.

43 Ibid.

44 Department of Justice, 'Domestic Violence Bill 2017 (as initiated)' (nd).

45 See n 41 above.

46 Crowley (n 42 above) 137.

47 Ibid.

48 Ibid 149.

49 See, not exhaustively, Women's Aid, 'Domestic Violence Bill 2017 Submission' (February 2017); Safe Ireland, 'Briefing for Members of the Oireachtas Legislative Amendments Recommended to: Domestic Violence Bill 2017' (April 2017); NWCI, 'Recommendations for Legislative Amendments: Domestic Violence Bill 2017' (February 2017).

parliamentary debates on the Bill evidence that introducing an offence of coercive control and determining the wording of such a provision were not straightforward tasks. It became immediately clear in parliamentary debates that there was a reticence by government officials to provide statutory definitions of domestic violence and coercive control. During the first Seanad debate, Minister Fitzgerald stated, 'given the complexity of relationships and the range of behaviours that could be considered coercive or controlling, it would be very problematic to define that in statute' and 'extremely difficult' to prove beyond reasonable doubt in a criminal law setting.[50] Similarly, Minister of State at the Department of Justice and Equality, David Stanton, showed sympathy to the possibility of defining coercive control in the statute, yet he also expressed his concerns. Similarly to Minister Fitzgerald, he displayed his hesitation in relation to challenges posed by enforceability of the offence and noted a potential risk of 'counterclaims being made by perpetrators to undermine the victim's case'.[51] Minister Stanton also objected to the inclusion of a list of behaviours to guide judicial decision-making, in place of a definition, positing that it would run the risk of narrowing the courts' broad discretion to decide on domestic violence cases.[52] This, he argued, would limit courts' independence and restrict their capacity to determine what is relevant in a case.[53]

On the other hand, Senator Colette Kelleher highlighted at the outset of the first Seanad debate the need to include 'a clear and comprehensive definition of what constitutes domestic violence', encompassing non-physical forms of violence, to account for the prevalence of psychological abuse.[54] An implicit recognition was made that coercive control was being left without legal remedy, with reference to organisations' concerns that even getting civil protection orders for non-physical forms of violence was difficult.[55] Senator Alice-Mary Higgins stated that arguments against the introduction of an offence of coercive control went 'against the advice we have had from all of the NGOs working in this area'.[56] She suggested, instead, that the creation of such an offence would provide an additional layer of protection under the law, in the form of civil orders and criminal prosecution, prior to escalation in danger and (physical) violence. She stated, 'it simply provides another thread or strand, and is something that

50 See n 41 above.
51 Seanad Deb 4 July 2017, vol 252, no 11.
52 Ibid.
53 Seanad Deb 28 November 2017, vol 253, no 9.
54 See n 41 above.
55 Ibid; Women's Aid, 'Domestic Violence Bill 2017 Submission' (February 2017) 19.
56 See n 51 above.

moves beyond those individual instances to a pattern of behaviour'.[57] Senator Higgins recognised the need to 'move past' a perception that only physical abuse amounts to domestic violence, and to bring to the attention of practitioners the behaviours they should be looking for when assessing whether abuse occurs.[58]

The inclusion of a definition of domestic violence was rightly perceived to be a central aspect in precipitating a cultural shift. The Law Society recommended its addition, citing the need to ensure the Bill's compliance with the recognition of domestic violence in law and practice, as defined by article 3 of the Istanbul Convention.[59] Senator Ivana Bacik further noted that the Law Society argued the definition was not only desirable but also 'required' under the Istanbul Convention.[60] A number of senators, including Senator David Norris and Senator Colette Kelleher pushed for the inclusion of this definition to include coercive control.[61] It was argued that, instead of serving the victims/survivors, the current broad judicial discretion, facilitated by a lacuna with regards to a definition of domestic violence or guidelines, provided a fundamental issue of 'inconsistency and divergence of practice' in domestic violence cases.[62] Bacik problematised this divergence, and highlighted that there seemed to be a misguided belief by some legal practitioners, including judges, that there was an unwritten threshold to be satisfied in practice for a civil law order to be made.[63] The suggestion that some legal practitioners were proceeding on the basis that this threshold existed speaks volumes in relation to the utmost need to adopt statutory guidance in relation to domestic violence, in order to avoid miscarriages of justice.

Eventually, during the Seanad session of 28 November 2017, Minister of State David Stanton announced that the Bill would include coercive control as an enumerated criminal offence. The reason for the Government's shift from an unwillingness to enact an offence of

57 Ibid.
58 Ibid.
59 Ibid.
60 Ibid.
61 See n 41 above.
62 See n 51 above.
63 Ibid. Senator Bacik, speaking of structured statutory guidelines stated: 'This is particularly important when we are talking about orders such as safety orders, protection orders and barring orders because we know there is a divergence in practice in this regard and that district judges and practitioners refer colloquially to applicants having to reach a bar or threshold before they will satisfy the judge that an order may be made. There is no such bar or threshold. I remember people talking about this when I was in practice. I understand this misleading expression is still being used now; yet judges do not have available to them any statutory criteria to guide them.'

coercive control during the July 2017 debate to tabling an amendment involving the addition of a coercive control offence before the following debate in November 2017 is not evident. It appears, however, to be a result of a joint effort by civil society organisations, a number of senators, and the office of the Minister of State with the Department and the Office of the Attorney General.[64] The important role played by the advocacy and lobbying of Irish women's rights organisations – including Safe Ireland, the National Women's Council of Ireland (NWCI) and Women's Aid – is discernible in their submissions made during the legislative amendments process.[65] For instance, NWCI had insisted that the criminal law approach to domestic violence did not 'reflect the true experience of long-term domestic abuse, including coercive control' and argued that a specific offence would improve access to justice and effective prosecution.[66] Similarly, Safe Ireland had recommended adding a new offence of controlling and coercive behaviour, potentially following the model of the England & Wales' provision.[67] Despite the coercive control addition, the DVA 2018 as enacted did not end up defining what constitutes domestic violence or what amounts to coercive control.

Undoubtedly, the DVA 2018 brought forward other significant legislative advancements in relation to women's protection and access to effective remedies, such as broadening the scope of legal protection to non-cohabiting partners and introducing emergency barring orders, from which coercive control survivors may benefit.[68] The parliamentary debates on the DVA 2018 evidenced the considerable reforms still needed in the justice system to effectively provide remedies to domestic violence survivors. Some concerns raised through the bill's various stages mirror issues present in other jurisdictions, such as the role that the judicial system plays in functioning as an extension of controlling behaviours by abusive partners or the current inadequacy of training for legal personnel and its practical negative ramifications for victims/survivors.

In the next section, I explore in more detail the merits and potential shortcomings of the coercive control offence in the DVA 2018 and questions which arise regarding its practical implementation.

64 See n 53 above.
65 Safe Ireland (n 49 above); NWCI (n 49 above); Women's Aid (n 49 above).
66 NWCI (n 49 above) 2.
67 Safe Ireland (n 49 above) 13.
68 See n 41 above.

THE CODIFICATION OF COERCIVE CONTROL IN IRELAND: ACHIEVEMENTS AND LIMITS

The offence of coercive control, as ultimately enacted in section 39 of the DVA 2018, reads as follows:

> (1) A person commits an offence where he or she knowingly and persistently engages in behaviour that— (a) is controlling or coercive, (b) has a serious effect on a relevant person, and (c) a reasonable person would consider likely to have a serious effect on a relevant person.
>
> (2) For the purposes of subsection (1), a person's behaviour has a serious effect on a relevant person if the behaviour causes the relevant person— (a) to fear that violence will be used against him or her, or (b) serious alarm or distress that has a substantial adverse impact on his or her usual day-to-day activities.
>
> (3) A person who commits an offence under subsection (1) is liable— (a) on summary conviction, to a class A fine or imprisonment for a term not exceeding 12 months, or both, and (b) on conviction on indictment, to a fine or imprisonment for a term not exceeding 5 years, or both.
>
> (4) In this section, a person is a 'relevant person' in respect of another person if he or she— (a) is the spouse or civil partner of that other person, or (b) is not the spouse or civil partner of that other person and is not related to that other person within a prohibited degree of relationship but is or was in an intimate relationship with that other person.

Several key legal elements are noteworthy. The Act sets out the relevant *mens rea* requirement as that of 'knowingly and persistently' engaging in the impugned harmful behaviour. The second requirement of 'serious effect' defines the threshold of the offence against the impact it has on the victim/survivor: either provoking fear *or* leading to day-to-day activities being substantially impacted due to a serious alarm or distress provoked by the behaviour. A 'serious effect' must be considered *likely* to arise as a result of the perpetrator's behaviour by a *reasonable person*. The 'relevant person' requirement defines the application of the offence to partners and ex-partners, regardless of the legal status of their intimate relationship.

The enactment of the coercive control offence is in itself an achievement as regards implementing article 33 of the Istanbul Convention and providing victims/survivors of coercive control a clearer legal avenue for redress. The criminalisation of coercive control not only allows criminal prosecution but also provides a springboard for civil remedies such as barring orders. On paper, the normative context of the Irish offence fully complies with the state's obligation contained in article 33. The Convention drafters provided states with leeway to decide how to legally define 'intentional conduct', which in

section 39 of the DVA 2018 appears as the requirement of 'knowingly' and 'persistently' engaging in behaviour that is considered as coercive control. Similarly, the requirement to affect the victim/survivor substantially, so as to seriously impair and damage their psychological integrity (found in art 33 of the Convention) is reflected in the DVA 2018's requirement that the behaviour has a 'serious effect' on the victim/survivor. This effect must be either fear of violence or serious alarm and distress affecting her on a day-to-day basis.

Despite the lack of definition, the listing under section 5 of the DVA 2018 of the factors and circumstances that judges must consider when determining applications for civil law orders can be seen as an achievement in relation to coercive control. The non-exhaustive list of contextual factors to be considered include: animal cruelty; a recent separation between the perpetrator and survivor; any deterioration in the survivor's physical or emotional wellbeing; any economic dependency; previous history of violence; and convictions for an offence under the Criminal Justice (Theft and Fraud Offences Act) 2001, or for an offence involving violence or threat thereof against any person.[69] The statutory addition of these factors is a welcome clarification as to how judges and practitioners may expose the presence, as a matter of law, of coercive control within a relationship.

However, shortcomings in the normative content are also noteworthy. Ireland missed the opportunity to be the first jurisdiction to adopt a statutory provision that addressed the gender power dynamics present in coercive control. Stark has argued that legal recognition of coercive control as a form of harm should go beyond 'adding new offensive behaviours to a series of (already unenforced) distinct offences'.[70] A gendered provision would highlight coercive control as a 'singular malevolent intent to dominate', which is 'most prevalent and has its most devastating consequences in heterosexual relationships', where the perpetrator's gender and male privilege justifies its use to enforce female subordination in a patriarchal society.[71] It is posited that the failure to do so is reflective of a lack of political will to be precise and clear in legally recognising the issue as a gendered one, unfortunately not unique to Ireland. All statutory offences criminalising coercive control in Europe, including England & Wales, Scotland, and Northern Ireland are gender-neutral even though they are backed by a governmental understanding of domestic

69 Domestic Violence Act 2018, s 5(2)(a)–(r).

70 Evan Stark, 'The "coercive control framework": making law work for women' in McMahon Paul (n 2 above) 40.

71 Ibid.

violence as a gendered phenomenon.[72] A codification of coercive control recognising its gendered power dynamics could have oriented the judiciary's implementation of coercive control in a gender-sensitive manner.[73]

Moreover, Ireland decided to adopt almost identical wording to section 76 of the Serious Crime Act 2015, now amended in the Domestic Abuse Act 2021, which applies in England & Wales, foregoing the opportunity to align with an alternative construction, such as Scotland's, which is considered 'one of the most radical attempts yet to align the criminal justice response with a contemporary feminist conceptual understanding of domestic abuse as a form of coercive control'.[74] This may be understood in light of the timing, given that stages of the Scottish parliamentary debates took place slightly later than Ireland's, perhaps not providing very much space for cross-fertilisation.[75] The formulation used in England & Wales and Ireland places the onus on the victim/survivor to show that the coercive and controlling behaviour has had a particular effect on her, namely a serious one which impacts her usual day-to-day activities. Focusing on the perpetrator's intent to harm, as does section 1 of the Domestic Abuse (Scotland) Act 2018, rather than on the victim/survivor's reaction to such harm, would have been strongly preferable.

Although civil society organisations had lobbied for the inclusion of more stringent penalties, the legislation provides for a maximum of 12 months' incarceration on summary conviction and a maximum of five years on conviction on indictment.[76] This is worth noting, given that harassment charges under the Non-Fatal Offences Against the Person Act 1997 can carry a sentence of up to seven years in prison.[77]

Among the normative shortcomings of the DVA 2018 in relation to coercive control, the lack of definition, paired with the absence of

72 Domestic Violence Act 2018, s 5(2)(a)–(r) (Ireland); Michele Burman and Oona Brooks-Hay, 'Aligning policy and law? The creation of a domestic abuse offence incorporating coercive control' (2018) 18 Criminology and Criminal Justice 67; Dáil Deb 15 December 2017, vol 963, no 4; 'Domestic Abuse and Family Proceedings Bill, Explanatory and Financial Memorandum' (nd). Even though there is no explicit recognition that coercive control is a gendered phenomenon in the Northern Ireland Domestic Abuse and Family Proceedings Bill Explanatory and Financial Memorandum, it may be argued that the document implicitly recognises it by making reference to the Bill's compliance with the UK's obligations under the Istanbul Convention.

73 UN CEDAW Committee, General Recommendation No 33 on Women's Access to Justice (23 July 2015) para 14.

74 Evan Stark and Marianne Hester, 'Coercive control: update and review' (2019) 25 Violence Against Women 81, 85.

75 The Scottish Parliament, 'Domestic Abuse (Scotland Bill)' (nd).

76 See n 53 above.

77 Non-Fatal Offences Against the Person Act 1997 Number 26 of 1997 (Ireland).

policy guidelines, is the most significant missed opportunity. Domestic violence – and especially coercive control – is not a universally understood issue warranting no definition or guidance. Far from it, the absence of such will very likely contribute to an inconsistency in the implementation of the offence.[78] As a matter of fact, concerns surrounding divergence of understandings, or the misunderstanding of the offence, and the consequential legal implications of the same, have fuelled academic debate over whether the positives of criminalising coercive control can outweigh the potential negatives.[79] Thus, it is possible that the detailed statutory guidance of offences enshrined in neighbouring jurisdictions becomes a relevant tool as Irish legal practitioners seek clarification, especially given the extreme similarity of Ireland's provision with that of England & Wales.[80] Against the background of this crucial deficiency, I now turn to the consideration of the potential issues with the provision's implementation in Ireland.

THE QUESTION OF IMPLEMENTATION: THE GATEWAY TO EFFECTIVE REMEDIES

Any theoretical discussion about the Irish criminalisation of coercive control should be accompanied by a preliminary evaluation of its practical implementation and the factors that can impact survivors' access to effective remedies. As advanced above, under the DVA 2018, effective legal remedies for survivors might take the form of protection measures under civil law or prosecution of the perpetrator under criminal law.[81] At the time of writing, there have been only a handful of successful prosecutions under the coercive control offence. Yet,

78 Women's Aid (n 49 above) 9.

79 Vanessa Bettinson and Charlotte Bishop, 'Is the creation of a discrete offence of coercive control necessary to combat domestic violence?' (2015) 66 Northern Ireland Legal Quarterly 179; Heather Douglas, 'Legal systems abuse and coercive control' (2018) 18 Criminology and Criminal Justice 84; Sandra Walklate et al, 'Is more law the answer? Seeking justice for victims of intimate partner violence through the reform of legal categories' (2018) 18 Criminology and Criminal Justice 115; Sandra Walklate and Kate Fitz-Gibbon, 'The criminalisation of coercive control: the power of law?' (2019) 8 International Journal for Crime, Justice and Social Democracy 94.

80 Home Office, 'Controlling or Coercive Behaviour in an Intimate or Family Relationship: Statutory Guidance Framework' (December 2015); Domestic Abuse (Scotland) Act 2018, s 2.

81 For more on effective remedies, see UN CEDAW Committee (n 73 above) para 19(b): 'remedies should include, as appropriate, restitution (reinstatement); compensation (whether provided in the form of money, goods or services); and rehabilitation (medical and psychological care and other social services). Remedies for civil damages and criminal sanctions should not be mutually exclusive'.

reporting of the crime continues to grow. In August 2022, 289 crimes had been reported, with charges having been brought in 53 cases.[82]

In February 2020, just over a year after the DVA 2018 entered into force, the first conviction and sentencing under section 39 of the DVA 2018 were handed down in the Letterkenny Circuit Court.[83] The case in question involved a man who was sentenced to 21 months of prison after pleading guilty to an array of charges against his ex-girlfriend, including coercive control, harassment and threats to damage property.[84] Reports of the case recounted extreme harassment, with the perpetrator calling the victim/survivor close to six thousand times over four months, as well as using other controlling and threatening tactics.[85] In November 2020, exactly nine months after the first sentence was handed down, another man, who previous to the trial had pleaded not guilty, was convicted of coercive control and assault against his ex-partner in the Dublin Circuit Criminal Court. This was the first court case under the DVA 2018 judged by a jury. The perpetrator was sentenced to 10 years and six months in prison, a sentence length which reflects his charges for coercive control as well as for repeated physical attacks.[86] Since, coercive control sentences have followed in a few reported cases, including in June 2021, July 2022 and January 2023. The first case involved a pattern of coercive control, including threats, to which the perpetrator pleaded guilty and was sentenced to three years in prison.[87] In the second case, the perpetrator was charged with a sentence of three years and three months after pleading guilty to harassment, assault causing harm, criminal damage, threats to cause criminal damage, endangerment, theft and threats to kill.[88] The last case involved a perpetrator who was sentenced to five years in prison for two counts of assault causing harm and one count of coercive control. In imposing sentence, Judge Sheahan recognised that the survivor had

82 Connor Lally, 'Surge in coercive control cases reported to Garda last year' *Irish Times* (Dublin 1 August 2022).

83 An Garda Síochána, 'First conviction and sentencing for coercive control in Ireland' (11 February 2020).

84 Stephen Maguire, 'Man jailed for coercive control phoned woman 5757 times in four months' *Irish Times* (Dublin 11 February 2020)

85 Ibid.

86 Liz Dunphy and Brion Hoban, 'Landmark coercive control sentence a warning to all abusers – charity' *Irish Examiner* (Dublin 22 January 2021).

87 Ann Healy, 'Coercive control: conspiracy theorist's reign of terror over family' *Irish Examiner* (Dublin 8 June 2021).

88 Conor Gallagher, 'Garda jailed for coercive control of terminally ill partner over four-year period' *Irish Times* (Dublin 26 July 2022).

'endured great upset and trauma from the assaults and psychological injury'.[89]

At this time, it is still uncertain how adequately actors within the Irish justice system approach the coercive control offence. At the time of writing, almost three years after the entry into force of the DVA 2018, information on the practical implementation of the offence remains scarce or not in the public domain. It is reported that as of July 2022 there had been 53 charges of coercive control.[90] Yet, there is lack of precise data, for example in relation to the number of applications for civil protection orders and orders granted in relation to coercive control.[91] This is consistent with the generally poor collection of criminal and civil justice statistics in Ireland. A recent response by the Minister for Justice to a parliamentary question regarding the number of domestic violence reports to police within a certain period is reflective of this. Minister McEntee asserted that attempting to establish unique persons who reported domestic violence in Ireland during the period between October 2020 to January 2021 'would require the expenditure of a disproportionate amount of staff time and resources in order to provide suitably accurate figures'.[92] As published by the Central Statistics Office (CSO), recorded crime in Ireland is marked 'under reservation', highlighting that its quality does not meet CSO standards.[93] Data are not disaggregated by crime, gender or relationship, making it impossible to access the number of reported cases under the new offence. Without a doubt, improving data collection is a challenge which must be prioritised, as having a direct impact on further research and related policy developments.

Policing coercive control

For the effective implementation of the offence, the response of the police and its involvement is crucial, especially in the identification of victims/survivors.[94] An Garda Síochána, has a Divisional Protective Services Unit in all Garda divisions, which, as part of the Garda National Protective Services Bureau, specialises in investigating crimes such as

89 Eimear Dodd, 'Man (21) jailed for coercive control and assault of ex-partner who "endured great upset and trauma"' *Irish Times* (Dublin 20 January 2023).

90 Dáil Éireann Debate (5 July 2022).

91 Courts Service, *Annual Report 2021* (July 2022).

92 Dáil Éireann Questions (660, 661) (3 March 2021).

93 Central Statistics Office, 'Recorded Crime – Under Reservation' (nd).

94 Cassandra Wiener, 'Seeing what is "invisible in plain sight": policing coercive control' (2017) 56 Howard Journal of Crime and Justice 500; Sian Dickson, 'Court of Appeal: coercion, control and assault: the importance of proactive policing and judicial standards in s 76 prosecutions R v Conlon (Robert Joseph James)' (2018) 82 Journal of Criminal Law 123, 125.

domestic violence.[95] Monitoring their role in responding to coercive control as a new statutory offence will be crucial to determine whether the concerns raised regarding policing and identifications in other jurisdictions are also relevant in Ireland.[96]

Research conducted on the policing of coercive control in England & Wales reflects that the criminalisation of coercive control, seen as a cultural change, necessarily requires a cultural shift in the police force.[97] A cursory look at the information provided by An Garda on its website regarding coercive control is promising. The website provides detailed information on the signs of coercive control, both for victims/survivors and relatives, and explains the process of reporting coercive control to the police. It advertises that Gardaí *will* provide advice, seek the victim/survivor's statement, initiate an investigation, gather evidence and witness statements, and examine tech belongings, such as phones or computers, in order to submit a file to the Office of the Director of Public Prosecutions.[98]

More detailed information on the policy guiding An Garda's intervention in domestic abuse cases can be found in its 2017 Domestic Abuse Intervention (DVI) policy, developed in consultation with statutory bodies such as the Office of the Director of Public Prosecutions and COSC (the National Office for the Prevention of Domestic, Sexual and Gender-based Violence).[99] The DVI pre-dates the DVA 2018 but nevertheless provides a reasonably comprehensive account of how intervention in domestic violence cases should be carried out. It helpfully includes a broad definition of domestic abuse and sets out good practice standards. The policy directs officers to, *inter alia*, take note of the history of abuse and current risk, note the physical and emotional condition of all parties, and use the power of arrest regardless of the victim/survivor's 'attitude' to it. It also takes into account Ireland's multicultural society and has a section that specifies cultural issues which may arise in the course of police intervention.

---

95 Department of Justice, 'Minister McEntee welcomes completion of rollout of Garda Divisional Protective Services Units' (29 September 2020).

96 See eg Charlotte Bishop and Vanessa Bettinson, 'Evidencing domestic violence, including behaviour that falls under the new offence of "controlling or coercive behaviour"' (2018) 22 International Journal of Evidence and Proof 3; Julia R Tolmie, 'Coercive control: to criminalize or not to criminalize?' (2018) 18 Criminology and Criminal Justice 50; Walklate et al (n 79 above); Charlotte Barlow et al, 'Putting coercive control into practice: problems and possibilities' (2019) 60 British Journal of Criminology 160.

97 Wiener (n 94 above) 503; Evan Stark, 'Looking beyond domestic violence: policing coercive control' (2012) 12 Journal of Police Crisis Negotiations 199, 213.

98 An Garda Síochána, 'Domestic abuse' (nd).

99 An Garda Síochána, 'Domestic Abuse Intervention Policy' (2017).

The DVI has been praised for its sensitivity regarding the 'understanding of the complexities of the abusive relationship and the cycle of abuse that typically occurs'.[100] Nevertheless, in light of the DVA 2018, an update would be most welcome to more comprehensively and specifically address particular dynamics of coercive control and highlight good intervention and investigation practices. As Wiener describes, police officers interviewed in London explained that investigating coercive control required shifting their approach, usually geared towards investigating concrete events and physical violence only, to one which included looking into less visible forms of abuse.[101] Thus, to appropriately identify coercive control when responding to a domestic violence call, police need to leave behind a stereotypical model that thinks of violence as equal to violent events.[102]

A further look into the Irish policing practice shows that considerable weight is given to 'incidents' of violence, and, in particular, to their seriousness. For example, in the DVI, An Garda Síochána indicates that 'the scale of abuse in previous incidents' should be considered during their intervention. Such a focus on the scale of the abuse could lead to interventions that undermine survivors' experiences, failing to recognise the continuums of abuse typical of the offence of coercive control. Whilst physical violence usually features very clearly delineated specific incidents with 'a degree of specificity (in time and space)', investigating coercive control requires a deeper look into relationship context and dynamics, which is 'more complex and time-consuming'.[103] Thus, in investigating coercive control, which is a course of conduct rather than incidental, investigators must pay special attention to the context in which the abuse takes place. Contextualising the abuse provides meaning to behaviours that may not seem harmful in isolation and often unveils a power dynamic that may not be immediately apparent, but that is nevertheless central to recognising the existence of the harm.[104]

It is true that the investigation of coercive control may bear similarities to the investigation of other offences which are not necessarily evidenced by visible external damage, for example, sexual assault or harassment. Generally, Tom O'Malley states in a recent report that training is required for all personnel coming into contact with victims under the European Union (EU) Directive on Victim's

100 Crowley (n 42 above) 151.
101 Wiener (n 94 above) 505.
102 Ibid 503.
103 Ibid 511, 512.
104 Bishop and Bettinson (n 96 above) 8.

Rights.[105] However, specific and specialised coercive control training, and indeed cross-cutting GBV training, will also be key for adequately investigating and identifying this form of harm. Data in England & Wales regarding the implementation of its coercive control offence evidences that coercive control is often misidentified and incorrectly labelled.[106] Tolmie explains that coercive control may be formed of behaviours that are not 'automatically unacceptable', and, therefore, the acceptability of these may 'depend on whether they are agreed to, and agreement can be the result of a matrix of factors'.[107] The literature provides many examples of this.[108] For instance, whereas at first glance asking one's partner to call when she gets to a place to ensure that she has arrived safely may seem caring, an unwritten ritual that entails punishment where she fails to do so may only be revealed if one pays attention to the nature and context, as well as the power dynamics of the relationship.

An Garda Síochána Commissioner, Drew Harris, has pledged to address coercive control, undertaking in-depth and effective investigations to gather evidence to support prosecution and to ensure accountability.[109] Reports detail that members of the Irish police force have received specific training and awareness on coercive control and the DVA 2018 at the national and frontline levels.[110] Moreover, An Garda Síochána announced in November 2021 a major review of its approach to domestic violence, including in light of international best practice.[111] Further training and a change of approach will be crucial in the implementation of coercive control as recent research suggests that cultural issues in policing domestic violence in Ireland remain, including 'problematic views of DV and abuse such as victim blaming, minimisation and patriarchal attitudes toward women', creating an 'inconsistent' response to domestic violence.[112]

105 Tom O'Malley, 'Review of Protections for Vulnerable Witnesses in the Investigation and Prosecution of Sexual Offences' (2020) para 10.6.

106 McGorrery and McMahon (n 6 above) 962.

107 Tolmie (n 96 above) 56.

108 Stark, *Coercive Control* (n 3 above).

109 Commissioner Drew Harris, 'Keynote Address' (Creating a Safer Ireland for Women: From Ratification to Implementation, Dublin, 6 December 2019).

110 Dáil Éireann Questions (795) (nd).

111 Conor Lally, 'Major Review will assess how Garda tackles domestic violence' *Irish Times* (Dublin 5 November 2021).

112 Stephanie Thompson et al, '"A welcome change … but early days": Irish service provider perspectives on Domestic Abuse and the Domestic Violence Act 2018' (2022) 1 Criminology and Criminal Justice.

Evidencing and prosecuting the offence

Bishop and Bettinson's work on evidencing coercive control highlights the unique difficulties arising in evidencing and prosecuting such offences.[113] They argue that, in order to evidence coercive control, an in-depth understanding of the behaviours in context, including the application of a gendered analysis, is crucial in order to ascertain the presence of the harm and the several forms that it can adopt. GREVIO reports illustrate that there is a cross-cutting issue across state parties to the Istanbul Convention regarding a lack of prosecution of psychological violence due to poor understanding and recognition of this form of harm as legally punishable.[114] Training, not only on evidence-gathering but also on the broader social context in which the harm takes place, is thus indispensable for all actors within the justice process.[115] The Office of the Director of Public Prosecutions points to 'lack of evidence' as the 'most common reason for decisions not to prosecute'.[116] In this line, it is promising that members of the police force are reportedly receiving training in investigative interviewing specifically geared towards victims/survivors of violence including psychological abuse and including relevant topics such as vulnerability and trauma narratives.[117]

Not only must evidence be available, but it must also be 'admissible, relevant, credible and reliable' and enough to prove the perpetrator's guilt beyond reasonable doubt.[118] So far, only cases where it may have been relatively straightforward to evidence coercive control beyond a reasonable doubt have reached criminal convictions. As stated, the first case featured a man calling his partner over five thousand times in four months. The second subjected the victim/survivor to economic control, having absolute control over her finances, humiliating her in public and isolating her from friends and family. Arguably, these were

113 Bishop and Bettinson (n 96 above).

114 GREVIO, 'Baseline Evaluation Report on Legislative and Other Measures Giving Effect to the Provisions of the Council of Europe Convention on Preventing and Combating Violence against Women and Domestic Violence (Istanbul Convention) Albania' (27 November 2017); GREVIO, 'Belgium' (n 34 above); GREVIO, 'Baseline Evaluation Report on Legislative and Other Measures Giving Effect to the Provisions of the Council of Europe Convention on Preventing and Combating Violence against Women and Domestic Violence (Istanbul Convention) Serbia' (22 January 2020); GREVIO, 'Baseline Evaluation Report on Legislative and Other Measures Giving Effect to the Provisions of the Council of Europe Convention on Preventing Combating Violence against Women and Domestic Violence (Istanbul Convention) Spain' (25 November 2020).

115 Walklate et al (n 79 above) 121.

116 Office of the Director of Public Prosecutions, 'Decision to Prosecute' (nd).

117 Dáil Éireann Questions (795) (n 110 above).

118 Director of Public Prosecutions, *Guidelines for Prosecutors* 5th edn (2019) 13.

not particularly challenging cases to identify nor evidence. It remains to be seen if and how more complex cases of coercive control will be identified and prosecuted.

Lack of guidance and victim/survivor's experience

As the victim/survivor's 'experience-based testimony'[119] is crucial for the determination of coercive control,[120] both at the identification stage and throughout the criminal process, it is fundamental that legal professionals and the judiciary are aware of the dynamics and workings of coercive control to ensure non-revictimisation and accountability. Where a case presents unclear facts, or there is no apparent physical violence, it will be necessary to establish whether the survivor has suffered from fear of violence or serious alarm and distress that has a 'substantial adverse impact' on usual day-to-day activities.[121]

The important weight given to the victim/survivor's reaction to the harm, judged through a reasonable person test, could require that the victim/survivor had a *specific* and performative response to her abuse *and* that a reasonable person would have considered their reaction likely. Rightly, Burman and Brooks-Hay express their concern as follows:

> [d]rawing lessons from sexual offence trials, the likelihood of the strategic use of evidence challenging victim credibility and character and suggesting 'motive to lie' in such circumstances is high, with attendant implications from shifting the trial focus from the accused's actions to those of the victim.[122]

Legal practitioners must be very wary of allowing stereotypes to play a role in decision-making. Lack of knowledge or awareness of the impact of coercive control on victims/survivors creates the perfect scenario for stereotyping regarding who constitutes a victim/survivor, based on both individual and 'larger discriminatory structures'.[123] For example, complainants refusing to make statements or later retracting them are recognised scenarios as regards domestic violence and coercive control.[124] These behaviours can be explained through victims/survivors' fear or allegiance to the perpetrator.[125] Nevertheless, in the absence of an adequate understanding of the manifestations and consequences of the offence, as well as victims/survivors' reaction to

119 Stark and Hester (n 74 above) 87.
120 Walklate et al (n 79 above) 119.
121 Domestic Violence Act 2018, s 39(2)(b).
122 Burman and Brooks-Hay (n 72 above).
123 Stark and Hester (n 74 above) 241.
124 Dickson (n 94 above) 123.
125 Ibid 125.

it, it can be easy to discredit their experiences as false, unconvincing or undeserving of protection.

Hence, effective identification, recognition and responses to coercive control victims/survivors will require professionals to challenge their preconceived ideas of what amounts to domestic violence and how victims/survivors should react to it. This will require a gender-sensitive analysis that is both mindful of how socially normalised gender dynamics may be weaponised to constrain victims/survivors' agency and that understands victimhood and remedy-seeking beyond a victim/non-victim dichotomy. As Hanna explains, 'the law forces the question of illegal coercion into a yes or no answer', and 'you are either coerced or not'.[126] For an effective operationalisation of the offence, legal practitioners will have to be willing to approach the 'serious impact' requirement in light of victims/survivors' ongoing struggle between resistance and victimisation, and how structural barriers such as preconceived notions of gender, race and victimhood, including their intersections, serve to make victims/survivors' victimisation unique.[127]

Structural issues

Some of the concerns relating to the implementation of coercive control legislation reflect issues with how the broader legal system interacts with victims/survivors of GBV. As introduced above, and focusing on domestic violence, Safe Ireland has submitted that fragmented, inconsistent responses to domestic violence victims/survivors continue and that 'research has highlighted systematic failings to implement current protections in legislation and policy'.[128] Even prior to the entry into force of the coercive control offence, women's rights organisations highlighted women's experience of abuse and the legal system, including women not being heard or being affected by stereotypes

126 Cheryl Hanna, 'The paradox of progress: translating Evan Stark's coercive control into legal doctrine for abused women' (2009) 15 Violence Against Women 1458, 1468.

127 Khatidja Chantler, 'Independence, dependency and interdependence: struggles and resistances of minoritized women within and on leaving violent relationships' (2006) 82 Feminist Review 27. See further Home Office, 'Review of the Controlling or Coercive Behaviour Offence' (March 2021). Citing Wiener (forthcoming): 'based on the qualitative research she conducted with judges, it penalises resilience in victims – the more able a victim is to withstand the controlling or coercive tactics of their partner, the lower the chances are that the requirement to prove adverse effect will be met'.

128 Safe Ireland, 'Changing Culture and Transforming the Response of Gender-Based Violence in Ireland. Submission to the National Women's Strategy 2017 – 2020' (January 2017); Safe Ireland, 'Department of Justice Criminal Justice Strategy. Submission from Safe Ireland' (10 August 2020).

based on 'gender, ethnicity, perceived class or level of education'.[129] These concerns, pre-dating the DVA 2018, aggravate the uncertainties around the effective implementation of the coercive control offence.[130] Moreover, so far it is unclear how the Irish justice system will address the concern that perpetrators use the legal process as a way to extend the abuse and further control victims/survivors, or indeed if such risk can be sidestepped.[131] Even though the weaponisation of the justice system by perpetrators is not a consequence of the DVA 2018, it is of particular relevance to victims/survivors of coercive control, as the justice system, which is lengthy and costly, has been recognised as a dangerous form of retaining coercion and control beyond separation.[132]

Another issue that further aggravates victim/survivor's negative experience with the justice system is the fragmentation between the criminal and civil legal systems.[133] A woman seeking a civil law order, or involved in child custody proceedings, will have to resort to the family system under civil law. In parallel, she may also have to attend criminal court as a witness where the perpetrator has breached a protective order or where the abuser is being prosecuted for their criminal behaviour. The disconnect between the two results in, *inter alia*, a lack of information-sharing between courts, excessively lengthy proceedings, or multiplicity of court proceedings.[134] The victims/survivors have to navigate through a complex system as it is, with an offence that – as it has been suggested – 'may require a breadth of evidence and complexity of analysis that the … system is not currently well equipped to provide'.[135] The impact and re-victimisation victims/survivors may suffer due to the structural issues present in the justice system, including steering through the two systems, undoubtedly requires further scrutiny.

---

129 Safe Ireland, 'The Lawlessness of the Home – Women's Experiences of Seeking Legal Remedies to Domestic Violence and Abuse in the Irish Legal System' (2014) 53; Women's Aid and Mazzone (n 1 above).

130 Criminal Justice (Victims of Crime) Act 2017 (Ireland). Transposing the EU Directive 2012/29/EU, the 2017 Act gives victims enhanced protection including the possibility of obtaining special protection measures to avoid repeat victimisation. It remains to be seen whether the Act has a significant impact on victims of coercive control as they interact with the justice system.

131 Douglas (n 79 above).

132 Ibid 86.

133 Safe Ireland, 'The Lawlessness of the Home' (n 129 above).

134 Women's Aid and Mazzone (n 1 above).

135 Tolmie (n 96 above).

CONCLUSION

The criminalisation of coercive control in Ireland is to be commended, as it addresses the historical legal isolation of specific incidents of abuse and validates the victims/survivors' lived experiences of psychological abuse.[136] Legislative reform brought by the DVA 2018 also constitutes a step towards implementing the comprehensive human rights standards of the Istanbul Convention.

However, this article has raised a number of shortcomings in the substantive content of the law, as well as prevalent and new procedural issues, which may affect the implementation of the offence and impact victims/survivors in different, still unknown ways. Statutory guidance engaging with the procedural and structural challenges arising, and how to address them, will be essential to pave the way towards a fair, equal and effective implementation of the coercive control offence. As Conaghan has rightly argued, the operation of the law, enforced through people and institutions, is intimately linked to 'deeply engrained, often unconsciously held social attitudes'.[137]

Cultural change in Ireland is required, as acknowledged in parliamentary debates, in light of Safe Ireland's recent research showing that archaic beliefs in terms of gender roles in society and victim-blaming culture are still widespread.[138] For the time being and in the absence of statutory guidance, if the coercive control offence is to have a positive impact on remedy-seeking for victims/survivors, it will have to be implemented in an intersectionally sensitive manner to account for victims/survivors' unique experiences, and in line with specialist guidance and training. This will require addressing, 'patriarchal attitudes and stereotypes, inequality in the family and the neglect or denial of women's civil, political, economic, social and cultural rights' and promoting women's empowerment.[139]

---

136 Ibid 51; Department of Justice and Equality, 'Statement by Minister Flanagan on Domestic Violence' (11 December 2019).

137 Joanne Conaghan, 'Some reflections on law and gender in modern Ireland' (2019) 27 Feminist Legal Studies 333, 335.

138 Safe Ireland, 'Gender Matters: Summary Findings of Research on Public Attitudes to Gender Equality and Roles, Domestic Abuse and Coercive Control in Ireland' (2019).

139 UN CEDAW Committee, 'General Recommendation No 35' (n 20 above) 34; Similarly, the latest CEDAW Committee periodic review of Ireland in 2017 highlighted that 'discriminatory stereotypes concerning the roles and responsibilities of women and men in the family and in society persist'. CEDAW, 'Concluding Observations on the Combined Sixth and Seventh Periodic Reports of Ireland' (9 March 2017) para 24(a).

Northern Ireland
Legal Quarterly
Winter Vol. 73 No. 4 (2022) 768–781
Commentaries and Notes DOI: 10.53386/nilq.v73i4.1062

R v Andrewes: judgment day for CV fraudsters? Case commentary on the Supreme Court decision reported at [2022] UKSC 24

Eli Baxter
University of Cambridge

Sarah Hair
Queen's University Belfast*
Correspondence email: ejsb2@cam.ac.uk; shair01@qub.ac.uk

ABSTRACT

Crime pays. Therefore, it is paramount that offenders are not permitted to retain the illicit profits derived from their course of offending. That is the purpose of the criminal law confiscation regime, which applies to a plethora of different offences where the state confiscates the ill-gotten gains the offender has retained after sentencing. This commentary focuses on one of these offences, the colloquially named *curriculum vitae* (CV) fraud. A relatively novel phenomenon in English law, CV fraud has come to the fore as a result of *R v Andrewes*, a recent Supreme Court decision. This commentary assesses this decision and ultimately concludes that while the Supreme Court's approach is sound in principle, it does not provide a solution which encompasses the broader spectrum of cases falling within the category of CV fraud. The *Andrewes* approach to the calculation of 'criminal benefit' may therefore require considerable adaptation in future cases. Perhaps most importantly, the absence of discussion on causation leaves this corner of the confiscation regime a grey area. This commentary sets out to offer a principled solution which might resolve this issue.

Keywords: *Andrewes*; proceeds of crime; confiscation; employment; CV fraud; causation; proportionality.

INTRODUCTION

The phenomenon of *curriculum vitae* (CV) fraud has recently found itself thrust into the legal limelight as a result of prominent litigation, news coverage, and academic discussion on the matter. In its simplest form, CV fraud denotes falsifying the details of one's educational history or work experience for the purposes of

* The authors extend their profound thanks to Dr John Stannard, Dr David Capper and John Larkin KC, whose insightful advice and guidance were indispensable in writing this article. Any remaining errors belong to the authors.

obtaining employment. The precise nature of CV fraud can vary in scope, ranging from a seemingly innocuous exaggeration as to one's skills and extracurricular interests, to the more serious cases which involve a consistent chain of falsehoods relating to the applicant's qualifications, academic achievements and employment history. It is on the more serious end of this spectrum where *R v Andrewes* lies, the final appeal in a course of litigation which provided the Supreme Court with the opportunity to determine the circumstances in which a confiscation order based on salary obtained through CV fraud will be proportionate.[1]

BACKGROUND

Mr Jon Andrewes applied for the position of Chief Executive Officer (CEO) at St Margaret's Hospice, Taunton, in October 2004. Under 'essential' requirements, applicants were to possess a first degree and 10 years' management experience with three years in a senior appointment. 'Desirable' attributes included an MBA and five years' experience in a senior appointment. In his application, Mr Andrewes claimed to hold a degree in social policy and politics, an MPhil in poverty and social justice, and an MBA in management science. He also claimed to be undertaking a PhD in ethics and management. Regarding his employment history, he indicated that he had been on secondment at the Home Office between 1979 and 1982, and had held numerous senior management and executive positions in the charitable sector from 1985 onwards.

However, these claims about his employment and educational history were no more than a 'staggering series of lies', which went undetected and enabled him to acquire the hospice CEO position in December 2004 at an initial annual salary of £75,000. Mr Andrewes maintained this façade during his tenure, and in 2006 informed his colleagues that he had completed his PhD, thereafter asking to be addressed as 'Dr Andrewes'. Using corresponding falsehoods, Mr Andrewes subsequently applied for and obtained two further remunerated appointments: first, the position of non-executive director at Torbay NHS Care Trust in 2007; second, chair of the Royal Cornwall NHS Hospital Trust in 2015. Despite the drastic disparity between his falsified background and actual experience, Mr Andrewes' performance was always appraised as either 'strong' or 'outstanding'. In 2015, however, the truth began to emerge, and Mr Andrewes' precariously assembled house of cards collapsed, bringing his employment and appointments to an end.

1 [2022] UKSC 24. *Andrewes* has since been followed in *R v Jiang (Shunjian)* [2022] EWCA Crim 1516, where the entire 'benefit' of an illegal enterprise was confiscated.

PROCEDURAL HISTORY

Criminal proceedings were initiated on three counts: one count of obtaining a pecuniary advantage by deception under section 16(1) of the Theft Act 1968; and two counts of fraud by false representation under section 2 of the Fraud Act 2006. Mr Andrewes pleaded guilty to all three counts, and was sentenced to two years' imprisonment in March 2017, with His Honour Judge Mercer QC noting that Mr Andrewes had proliferated 'a series of staggering lies' on which his 'outwardly prestigious life' was precariously perched for over a decade.

Subsequent confiscation proceedings were heard by Recorder Meeke QC, who determined that Mr Andrewes' full earnings of £643,602.91 (net of tax and national insurance) constituted benefit from particular criminal conduct. The recoverable amount was £96,737.24,[2] and a confiscation order was made for that sum. The recorder rejected the submission that Mr Andrewes had not benefited from criminal conduct because he had earned remuneration from the work performed, and that any benefit was thus too remote. He also did not regard the confiscation order as disproportionate under the Proceeds of Crime Act (POCA) 2002,[3] saying that it represented less than 15 per cent of the benefit figure.

Mr Andrewes appealed against the confiscation order, and the Court of Appeal examined the case from the angles of causation and proportionality. The court found Mr Andrewes' causation arguments to 'fail at every level', concluding that provision of lawful and full value service for remuneration received did not break the causal chain.[4] The first reason given for this conclusion was Mr Andrewes' guilty plea to the three counts brought against him: 'by his pleas to counts 1–3 ... he plainly accepts, in terms, that he had [benefited from his particular conduct]'.[5] The court dismissed remoteness as a non-issue because the false representations were continuing[6] and cited the broadness of the

2 The recoverable amount is the lower figure of either the criminal benefit or the available amount, per POCA, s 7. In this case £96,737.24 was the available amount, 'available' meaning the amount Mr Andrewes could actually realise.

3 Under POCA, s 6(5).

4 *R v Andrewes* [2020] EWCA Crim 1055, [38]–[40], [68].

5 Ibid [26]. This assertion that a guilty plea equates to agreement or acquiescence that earnings were not too remote and that the money earned was thus criminal benefit runs the risk of retrospectively putting words in the mouth of the accused.

6 Ibid [70]–[72]. Continuation of misrepresentation is perhaps better categorised under factual causation than legal remoteness, in that 'but for' the continuance of his misrepresentations Mr Andrewes would have lost the opportunity to earn.

language of section 76(4) and (5) of POCA[7] to justify focusing simply on satisfaction of the 'but for' test. However, the court concluded that a confiscation order for any amount would be disproportionate because Mr Andrewes had given full value for his earnings in rendering services to a high standard.

The Crown appealed to the Supreme Court solely on the grounds of proportionality, Mr Andrewes having conceded that he had relevantly benefited from his criminal conduct. The Crown continued to advocate for a 'take all' approach (confiscation of full net earnings) while Mr Andrewes endorsed the 'take nothing' approach adopted by the Court of Appeal. The certified question for the Supreme Court was as follows:

> Where a defendant obtains remuneration as a result of or in connection with an offence of fraud based upon the obtaining of employment by false representations or non-disclosure, in what circumstances (if any) will a confiscation order based on the wages earned be disproportionate within the terms of section 6(5) of the Proceeds of Crime Act 2002, or contrary to Article 1, Protocol 1 (A1P1) of the European Convention on Human Rights?

IN THE SUPREME COURT

The Supreme Court unanimously allowed the Crown's appeal. Lord Hodge and Lord Burrows delivered the main judgment, with which Lord Kitchen, Lord Hamblen and Lord Stephens agreed. Before turning to the decision, however, it is first necessary to consider (albeit briefly) a chronology of the confiscation regime leading up to *Andrewes*.

Before the seminal judgment in *R v Waya*,[8] which is considered below, the confiscation regime was without a keystone. The first version of the POCA was passed in 1995, itself an amending statute intended to remedy the significant practical differences spawned by its piecemeal predecessors. However, after the enactment of the Human Rights Act (HRA) in 1998, the statute had to be read in a way which gave effect to rights[9] under the European Convention on Human Rights (ECHR) and in light of this, any application of the POCA became rather problematic. In particular, the 1995 Act effectively removed the Crown Court's discretion in the making of a confiscation order if

---

7 POCA s 76(4) provides: 'A person benefits from conduct if he obtains property as a result of or in connection with the conduct.' S 76(5) provides: 'If a person obtains a pecuniary advantage as a result of or in connection with conduct, he is to be taken to obtain as a result of or in connection with the conduct a sum of money equal to the value of the pecuniary advantage.'

8 [2012] UKSC 51.

9 HRA, s 3.

it was successfully applied for by the prosecution and the statutory requirements were satisfied.[10] The court therefore had no discretion to mould a confiscation order to secure justice in each case, opening the door to grossly disproportionate results.

Initially, the position remained the same under the POCA 2002. Under this version (as enacted), the court was only required to determine the recoverable amount of money and make a confiscation order for that amount.[11] While the POCA 2002 was subsequently certified as Convention compliant,[12] the underlying question remained: would the application of POCA's rules for calculating the confiscation order amount to a contravention of Convention rights? As the certified question indicates, the relevant Convention right has always been article 1, protocol 1 (A1P1) of the ECHR, which provides for the peaceful enjoyment of property to the extent that it is not prohibited by law, the public interest, or the principles of international law. This provision requires, via the rule of fair balance, that there be a reasonable relationship of proportionality between the means employed by the state in depriving criminals of their property and the legitimate aims sought by that deprivation.[13]

Nevertheless, in the cases preceding *Waya*, there was little reference to this notion of proportionality. In *R v Rezvi*,[14] for instance, Lord Steyn, whilst acknowledging the legitimate aim of depriving criminals of the profits of their crime, strongly emphasised the importance of punishment and deterrence, an emphasis which was cited and applied in many subsequent cases.[15] Yet, in light of this interpretative obligation, the reference to these provisos was plainly incorrect, specifically because the regime was not intended to be penal or deterrent in effect. A confiscation order is not an additional fine or penalty; it simply strips the profits of crime.[16] The punishment received by the offender is to be contained in the sentence passed following a finding of guilt, and confiscation proceedings are entirely distinct from the sentencing process. While *R v May* subsequently recognised the importance of proportionality, noting that a failure to consider the proportionality proviso could lead to an 'oppressive' interference with Convention rights,[17] there yet remained no explicit expression of proportionality in the confiscation regime.

---

10 *Waya* (n 8 above) [4].
11 POCA, s 6(5).
12 Under the HRA, s 19.
13 *Jahn v Germany* (2006) 42 EHRR 1084, [93].
14 [2002] UKHL 1.
15 *Waya* (n 8 above) [2].
16 Ibid.
17 [2008] UKHL 28, [42].

This brings us to *Waya*, where the Supreme Court recognised that, in order to avoid infringing A1P1, each confiscation order had to be proportionate to the legislative objective of stripping the profits of crime. In this, section 6(5) of the POCA was read as subject to the qualification that the court could make a confiscation order for the recoverable amount 'except insofar as such an order would be disproportionate and thus a breach of Article 1, Protocol 1'.[18] On the facts of *Waya*, a proportionate confiscation order was therefore not one which focused on the value of Mr Waya's fraudulently obtained mortgage (he had misrepresented his earnings), but rather one which focused on the benefit obtained, the increase in the value of the property acquired through the loan.[19]

Prior to this qualification, disproportionate confiscation orders were prevented by reliance on the court's jurisdiction to avert an abuse of process. For instance, in *R v Shabir*,[20] the defendant's benefit was calculated as £464, but the Crown nevertheless sought a grossly disproportionate £400,000 confiscation order as a result of the circumstances in which he obtained the money. These proceedings were rightly stayed for an abuse of process. While the Court in *Waya* accepted that this decision was plainly correct,[21] it believed that the better analysis was one based on proportionality, as this more adequately appreciated the special relationship between the POCA and ECHR. After *Waya*, the POCA was amended to include an explicit reference to proportionality, with section 6(5) now providing that the court must make a confiscation order for the recoverable amount 'to the extent that it would not be disproportionate to require the defendant to pay the recoverable amount'.[22]

Perhaps of equal importance was *Waya*'s recognition that a defendant could potentially restore the benefit gained through their criminal conduct in ways other than monetary repayment – certain acts could be analogous to restoration.[23] An example of this was a defendant who, by deception, induced someone to trade with them, but otherwise gave full value for the lawful goods or services obtained. While such individuals clearly deserved punishment (some form of criminal sentence), the question of whether a confiscation order was proportionate was entirely distinct and required no small degree of circumspection.[24] Nonetheless, it is imperative to note that *Waya*

18 *Waya* (n 8 above) [16].
19 Ibid [41], [79].
20 [2008] EWCA Crim 1809.
21 *Waya* (n 8 above) [18].
22 Inserted by the Serious Crime Act 2015, sch 4, para 19.
23 *Waya* (n 8 above) [34].
24 Ibid.

explicitly left the door open concerning the proper application of proportionality (and causation) to confiscation proceedings involving CV fraud, deeming that such questions were best answered in an appeal where they directly arose.[25]

Given the dearth of case law on this issue in the decade that followed, *Andrewes* was precisely that appeal. The Supreme Court seized this opportunity to reframe *Waya* whilst revisiting certain aspects of the proportionality regime which had caused practical difficulties. Crucially, it was noted that the Court of Appeal was mistaken in drawing a distinction between the notion of proportionality in section 6(5) of the POCA and that in A1P1 of the ECHR.[26] Here the Supreme Court reaffirmed the traditional *Waya* analysis by recognising that the section 6(5) POCA proviso *already* embraced the ECHR notion of proportionality and had to be read in a manner which gave effect to A1P1.

The relationship between these two provisions was explained more recently in *Bank Mellat v HM Treasury (No 2)*,[27] a leading case on the notion of Strasbourg proportionality in United Kingdom domestic law, which laid down a four-part proportionality test. Using this analysis, the court explained how the confiscation regime already satisfies the first three tenets of ECHR proportionality: it harbours the legitimate aim of disgorging the profits of crime; it is a rational means of doing so; and the goal cannot be achieved less intrusively. The sole issue was the fourth step, which asks whether the measure is a proportionate means of achieving the legitimate aim. In the court's view, this was precisely what the proviso in section 6(5) already asked – namely whether the confiscation of the relevant sum was a proportionate means of removing the profits of crime. Thus, it is clear that this distinction (if there ever was such) is no longer in existence; and the two tests are, at least in this context, now coterminous. While this issue did not appear to manifestly influence the Court of Appeal's decision, given it did not regard confiscation to be in any sense proportionate, the clarification is nevertheless welcome because it streamlines the theory of proportionality underlying the regime, and thus expedites confiscation proceedings in the Crown Court.

Turning to the submissions before the Supreme Court, neither party's approach found favour with the bench. The 'take all' approach argued by the Crown was disproportionate because such indiscriminate confiscation did not recognise or reflect a deduction for the value of the services legally provided by Mr Andrewes. In this vein, the Supreme Court agreed with the Court of Appeal that such recovery

25 Ibid.

26 *Andrewes* (n 1 above) [38].

27 [2013] UKSC 39, [20], [74].

would constitute a 'double penalty',[28] and thus double punishment, something explicitly recognised to be 'disproportionate and wrong'.[29] On the other hand, Mr Andrewes' 'take nothing' approach was equally unsatisfactory: in the absence of any confiscation order, he would effectively be permitted to retain benefit from his fraud, an affront to the regime's very purpose.

Therefore, faced with two 'extreme' approaches, one too harsh and the other too lenient, the Supreme Court turned to a 'principled middle way' in seeking to confiscate only the 'profit' of the fraud. Just what this meant was explained through the lens of *R v Sale*.[30] In this case, the defendant Mr Sale was the sole director and shareholder of a company which obtained valuable contracts with Network Rail by bribing one of its managers. Of the £1.9 million turnover deriving from these contracts, the court assessed the profit (before tax) as being £200,000. While Mr Sale, like Mr Andrewes, had given full value for the benefit obtained, because the contracts had been performed legally, efficiently and at market price, the court nevertheless reasoned that the previous confiscation order (the entire turnover of £1.9 million) should be replaced by one which confiscated only the £200,000 profit, thereby removing the 'benefit' of Mr Sale's crime.[31]

For Mr Andrewes, his 'profit' was determined as the difference between his pre-fraud earnings of £54,361 in 2004 and his fraudulently obtained higher earnings of £75,000 from 2004 onwards. This difference equated to a 38 per cent increase in net earnings. Applying this increase to his full net earnings of £643,602.91 amounted to a sum of £244,569 – the true 'profit' gained from his fraudulent course of employment. To simplify matters, the court noted that this was not a complex accounting exercise, and while some evidential reasoning remained necessary, the issue was to be approached in a broad-brush manner to simplify the administration of justice in Crown Court confiscation hearings.[32] This broad-brush approach is best evidenced through the manner in which the court simply added Mr Andrewes' further higher earnings from his two subsequent appointments without compounding the percentage benefit,[33] which would have amounted to a greater 'benefit' figure. Nonetheless, since the profit figure vastly exceeded the recoverable amount of £96,737.24 it was plainly proportionate to confiscate that

28 *Andrewes* (n 1 above) [42].
29 *Waya* (n 8 above) [28]–[29], [33].
30 [2013] EWCA Crim 1306.
31 *Andrewes* (n 1 above) [31].
32 Ibid [48].
33 Ibid [51].

amount and thus Recorder Meeke QC's confiscation order was duly restored, albeit supported by very different reasoning.[34]

THE NEW HALFWAY HOUSE

CV fraudsters beware: *Andrewes* signifies that even the provision of full value for the benefit obtained will not necessarily spare them from a confiscation order, no matter the quality of their work or ultimate value to their employer. Put simply, if an individual commits CV fraud, it will be proportionate to confiscate the difference between their pre-fraud earnings and the higher earnings obtained following their fraud. This new test is commendable as it re-tethers the confiscation regime to its legislative purpose – removing the *profits* of crime. The Court of Appeal seemingly focused on the wrong question in asking whether any order would be proportionate, notwithstanding the amount. In failing to remove the benefits of Mr Andrewes' crime, despite recognising that he had 'relevantly benefitted' from his criminal conduct, the Court of Appeal perhaps lost sight of the underlying objective of the POCA[35] and placed too much weight on the provision of full value for the wages received, an oftentimes unquantifiable calculation.

The confirmation that the provision of full value does not automatically bar a confiscation order represents some derogation from precedent. A few months prior to *Andrewes*, the Court of Appeal in *R v Asplin* analysed the precedents and concluded that 'if full value has been given for the benefit received, it will be disproportionate to make a confiscation order'.[36] While it was conceded that the provision of anything less than full value – such as 'significant value' – would render a confiscation order proportionate,[37] it appears that the courts in these cases had become unduly distracted with notions of full value and restoration. While the 'full value' consideration is indubitably important because a confiscation order against a defendant who has restored some of the benefit obtained should reflect this to remain proportionate, it ought not to be the preeminent consideration.

Attaching too much significance to the provision of full value could permit the defendant to retain significant criminal benefit. The question is now whether the defendant has disgorged the 'true' benefit of their crime; cases which pre-date *Andrewes*, to the extent that they informed

34 Ibid [52], [57].

35 *Waya* (n 8 above), [27].

36 [2021] EWCA Crim 1313 [33]. Precedents considered included *May* (n 17 above); *R v Morgan* [2008] EWCA Crim 1323; *R v Sale* [2013] EWCA Crim 1306. The Court of Appeal judgment in *R v Andrewes* [2020] EWCA Crim 1055 was also considered at this juncture.

37 Ibid [50].

decisions on full value and proportionality in making a confiscation order, must now be treated with a great degree of circumspection in future cases where the defendant has made restoration, or performed acts regarded as analogous to restoration.

Importantly, the Supreme Court also provided guidance on the issue of where the services provided are illegal *per se*, such as practising medicine without a licence. This distinction is paramount. The Supreme Court dedicated much of its judgment to an analysis of previous cases on this issue, chief of which was *R v King*,[38] where the Court of Appeal actually removed the turnover figure as opposed to the profit because the entire business enterprise was founded on illegality, rather than merely being tainted by it. *Andrewes* endorsed this approach, holding that illegal services have no value the law should recognise,[39] and therefore confiscation to that end will not constitute double disgorgement as the provision of illegal labour cannot 'restore' value. The pertinent consideration is whether the provision of services is a criminal offence; a legal bar to appointment is not sufficient to render a confiscation order for the full turnover or earnings proportionate.[40]

Overall, the Supreme Court's reaffirmation is immensely instructive. Two years prior to *Andrewes*, the Law Commission had noted how 'the absence of an overt statement' as to the purpose of the POCA regime meant that its central objective in disgorging the proceeds of crime had been marred by other ancillary objectives, thereby confounding the regime, and unfairly impacting upon the assessment of proportionality in other comparable cases.[41] While the defendant will undoubtedly view the regime as deterrent and penal in effect,[42] the essence of the legislation nevertheless rests squarely on the notion of disgorging the proceeds of crime. *Andrewes*, therefore, arguably represents the 'overt statement' the Law Commission desired, but its clarificatory effect is perhaps limited to cases of CV fraud, and where services provided were lawful.

THE MISSING LINK: CAUSATION

The question of causation remains unresolved. Given that the appeal focused on proportionality – causation having been conceded – the Supreme Court was not afforded scope to discuss this in detail. Thus, the short conclusion is that the approach applied by the Court of Appeal

38 [2014] EWCA Crim 621.

39 *Andrewes* (n 1 above) [42].

40 Ibid [54].

41 Law Commission, *Confiscation of the Proceeds of Crime after Conviction* (Consultation Paper No 249 2020) [5.36]–[5.37].

42 *R v Harvey* [2015] UKSC 73, [55].

in *Andrewes* must stand. In this context, causation is satisfied if (i) 'but for' the dishonest statements the defendant would not have secured employment; and (ii) the false representations continued throughout the employment.[43] While the focus on simple 'but for' causation corresponds with the broad-brush approach endorsed by the Supreme Court,[44] the inadequate consideration afforded to legal causation or 'remoteness' leaves it unclear precisely when the benefit obtained will be too remote from the relevant criminal conduct. This is particularly problematic given that *Waya* acknowledges that cases of employment obtained by deception may raise 'difficult questions of causation … quite apart from any argument based upon disproportion'.[45]

Should it be concluded that property is obtained 'as the result of or in connection with' an offence where the defendant's lawful conduct far surpasses the illegal conduct in operative effect? Neill LJ considered this question in *R v King and Stockwell*: 'the question in each case is: was the deception an operative cause of the obtaining of the property?.[46] A test of operative cause would give legal causation its due place, considering the contribution the unlawful act made to the benefit obtained versus other factors which may render benefit too remote. Inverse analogy might be drawn to the doctrine of *novus actus interveniens*, in that the defendant's own lawful actions detract from the consequences of the unlawful actions.[47] Lord Bingham's analogy in *R v May* is illustrative:

> If (say) a defendant applies £10,000 of tainted money as a down-payment on a £250,000 house, legitimately borrowing the remainder, it cannot plausibly be said that he has obtained the house as a result of or in connection with the commission of his offence.[48]

Applying this reasoning, a defendant's lawful contributing conduct throughout the employment *could* result in the fraud ceasing to be legally operative, rendering the benefit too remote. This rationale is consistent with POCA section 8(2)(a), which provides that the court 'must take account of conduct occurring up to the time it makes its decision' in determining whether the defendant has benefited from the criminal conduct. Such reasoning is also not novel: in the unreported

---

43 *Andrewes* (n 1 above) [26].

44 Ibid [48].

45 *Waya* (n 8 above) [34].

46 [1987] QB 547, 553.

47 Take, for example, the employee whose employer decides to give him a promotion and pay rise for exceptional performance: although such income would not be received 'but for' the fraud, its receipt is conceptually even further removed than base salary.

48 *May* (n 17 above) [26].

case of *R v Lewis*,[49] the court held the defendant was paid because of the services she rendered during her employment as a teacher, not because of her false representation that she had a teacher's licence. The Court of Appeal likewise recognised in some employment cases that false representations might cease to have operative effect,[50] but there is unfortunately no clarification as to when this might occur.

The foregoing is not to suggest legal causation should be used as a bulwark against confiscation orders by rendering the benefit entirely remote from the criminal conduct; rather, it might operate in tandem with proportionality to create a balanced regime that strips only the profits of crime, and nothing further.

FUTURE APPLICATION

A number of different factual scenarios demonstrate the practical difficulties in employing the *Andrewes* formulation in future cases. Firstly and most obviously, how would the principled middle way apply if an individual falsified their CV to obtain their first employment? According to the *Andrewes* algorithm, criminal benefit equals the difference between the lower pre-fraud earnings and the higher earnings obtained in connection with fraud. In such a case where previous earnings equal zero, the difference and thus criminal benefit would be the full net earnings, despite restoration of value by the defendant. Yet *Andrewes* reaffirms the confiscation order ought to 'reflect a deduction for the value of the services rendered', else it would constitute 'double disgorgement' and would be disproportionate.[51]

Likewise, it is unclear how a proportionate confiscation order is to be determined where an individual falsifies their CV to obtain a job which pays *less* than their previous position. Such circumstances could arise if an individual has lost their job in a highly specialised industry and must act quickly to secure employment elsewhere, or where an individual commits fraud by non-disclosure after having lost a better-paying position for some reprehensible reason they wish to hide. Here the pre-fraud earnings could be much higher than those obtained post-fraud, meaning that the defendant has not 'benefited' from their fraud per the *Andrewes* algorithm. Notwithstanding, following the *Andrewes* conclusion on causation, such individuals' earnings would constitute benefit obtained in connection with criminal conduct, thus potentially warranting some confiscation.

---

49 (Somerset Assizes, January 1922); see *Russell on Crime* 12th edn (Sweet & Maxwell 1964) vol 2, 1186.

50 *Andrewes* (n 1 above) [75].

51 Ibid [41].

Accordingly, the 'principled middle way' may have a narrow application. In our limited proffered examples, the equation would require significant alteration or revision; for example, the court might (arbitrarily) adopt the national average salary to serve as a lower figure if the individual has no prior comparable earnings. For those cases which fall entirely outside the remit of the *Andrewes* algebra – such as those where the higher and lower values are inverted – the court may need to revert to the drawing board, or at least undertake major modification of the equation, to ensure proportionality.

Furthermore, the court did not answer its own question as to how the middle way would work where the defendant is paid an upfront sum for lawful services and there has not yet been restoration.[52] In these circumstances, the grounds for disgorgement might lie in the forestated causation theory because payment has been received immediately after the fraudulent act, without lawful work intervening as a prerequisite to receipt of wages. However, if the defendant has commenced work after the 'golden handshake', the confiscation order may again require a deduction to reflect restoration provided to ensure proportionality. Once more, this falls outside the *Andrewes* algorithm; the court might instead consider what percentage or proportion of the services promised has been completed and constitutes restoration.

Thus, although the Supreme Court helpfully clarified the general principles to adopt when determining a proportionate confiscation order in cases of CV fraud, the 'principled middle way' will only work where there is evidence of quantifiable lower earnings and higher earnings, and where there is a legal market for services provided.[53] As *Waya* foresaw, 'prosecuting authorities may need to reflect long and hard before deciding on confiscation proceedings',[54] as without clear evidence of antecedents as regards previous legal earnings, confiscation proceedings might quickly prove futile. Additionally, in many cases the criminal benefit will vastly exceed the recoverable amount, in practice negating the need for precise or complex accounting exercises, given that section 6(5) of POCA instructs the court to address the proportionality of confiscating the recoverable amount rather than the benefit obtained.[55]

CONCLUSION

The principled middle way adopted by the Supreme Court, at least *prima facie*, deals well with the requirement of proportionality. In striving to

52 Ibid [55].
53 Ibid [53].
54 *Waya* (n 8 above) [19].
55 *Andrewes* (n 1 above) [50].

confiscate only the profits of the fraud, the ruling refocuses attention on the confiscation regime's true purpose – stripping the profits of crime. It is a judgment to be respected for (i) its clarification of proportionality in cases of CV fraud; (ii) its confirmation of the coterminous relationship between Strasbourg and POCA proportionality; and (iii) its clarification of proportionality in cases where the services provided are illegal. That said, this approach is not a panacea; its straightforward application is limited to specific timelines of CV fraud, potentially leaving avenues of future appeal open where the facts deviate from the *Andrewes* matrix. Likewise, whilst the need to avoid complicating the administration of justice in Crown Court confiscation proceedings is certainly valid, the absence of a reasoned conclusion on causation is unfortunate, given the theoretical and potential practical uses this line of reasoning possesses.

Northern Ireland
Legal Quarterly
Winter Vol. 73 No. 4 (2022) 782–806
Commentaries and Notes DOI: 10.53386/nilq.v73i4.1074

The Union in court, Part 2: *Allister and others v Northern Ireland Secretary* [2022] NICA 15

Anurag Deb
Queen's University Belfast
Gary Simpson
Queen's University Belfast
Gabriel Tan
Wilson Solicitors*
Correspondence email: adeb01@qub.ac.uk

INTRODUCTION

The Ireland/Northern Ireland Protocol (Protocol) to the Withdrawal Agreement between the United Kingdom (UK) and the European Union (EU) has become, if anything, more politically divisive and polarising with time, rather than less. Even as this case is destined for the UK Supreme Court,[1] the House of Lords is considering (the House of Commons has already passed) a Bill to disapply large parts of the Protocol in domestic law,[2] resulting in criticism at the breach of international law brought on by such a step.[3] Nevertheless, the UK Government defends the Bill as necessary to 'uphold' the Belfast (Good Friday) Agreement 1998 (GFA),[4] which is of critical importance to all aspects of governance and peace in Northern Ireland. The Northern Ireland Assembly has yet to elect a Speaker after fresh elections in

* We pay tribute to the Right Honourable William David Trimble, Baron Trimble of Lisnagarvey, one of the appellants (and original applicants) in this case, who died on 25 July 2022. Lord Trimble was a towering figure in Northern Ireland and is perhaps best remembered as an architect of the Belfast (Good Friday) Agreement 1998, which brought an end to decades of violence and bloodshed. We are grateful to Professor Colm Ó Cinnéide and the anonymous reviewer for their thoughtful comments and helpful feedback. Any errors and shortcomings remain our own.

1 John Campbell, 'Northern Ireland Protocol: Supreme Court set to hear challenge' (*BBC News* 25 April 2022).

2 Lisa O'Carroll and Heather Stewart, 'Northern Ireland protocol bill passes Commons vote' (*The Guardian* 27 June 2022) .

3 Ronan Cormacain, 'Northern Ireland Protocol Bill: A Rule of Law Analysis of its Compliance with International Law' *(Bingham Centre for the Rule of Law* 17 June 2022).

4 Foreign, Commonwealth and Development Office, 'Foreign Secretary: bill will fix practical problems the Protocol has created in Northern Ireland' (27 June 2022).

May 2022, and is thus unable to function in any capacity, while the Northern Ireland Executive exists in a largely 'caretaker' capacity, with no First or deputy First Minister.[5] At the heart of this great unravelling lies the Protocol, or more accurately, what it has come to signify in Northern Ireland and UK politics. Any legal challenges around the Protocol, therefore, are highly anticipated matters.

This comment follows a comment which one of us wrote, concerning the first instance judgment in this case.[6] Consequently, the factual matrix is not rehearsed in any great detail in this comment. Rather, after setting out some general critiques, we explore each of the five grounds of appeal before concluding.

At the outset, it should be noted that *Allister* met with the same fate in the Northern Ireland Court of Appeal (NICA) as it had in the High Court – a dismissal of the various challenges – but for somewhat different reasons.

GENERAL POINTS

Before we delve into the judgments, it is important to set out two general critiques.

First, the NICA judgments are astonishingly long. Together, they number 601 paragraphs, and are 83.8 per cent longer than the judgment of the High Court, which was set out in 327 paragraphs. Moreover, the concurring judgment of McCloskey LJ in the NICA, at 302 paragraphs, constitutes 50.25 per cent of the total NICA judgment. This reflects the fact that McCloskey LJ rehearsed the factual matrix of the case in similar detail as the Lady Chief Justice, whose judgment for the NICA majority has greater precedential value. We question whether this combined length was necessary, as there are certain segments which expansively set out either well-trodden precedents or the appellants' evidence and may have been usefully condensed.[7]

Second, the manner in which the NICA appears to have determined the appeal is also somewhat concerning. As will become clear in our substantive analysis of the judgments, several grounds of appeal appear to have been at least determined *de novo*. The judgments rarely examine, or even advert to, any errors (whether or not substantiated) on the part of the first instance judge, Colton J, which could have given rise to an appeal at all. This is surprising, as generally, appeals

5 'NI Election 2022: Prime Minister to visit NI as DUP blocks assembly' (*BBC News* 13 May 2022).

6 Anurag Deb, 'The Union in court: *Allister and others' Application for Judicial Review* [2021] NIQB 64' (2022) 73(1) Northern Ireland Legal Quarterly 138.

7 For example, the majority's discussion on the justiciability of treaty-making and the delay in bringing the original proceedings, *Allister and others' application for judicial review* [2022] NICA 15, [31]–[58].

in the NICA are conducted by way of re-hearing rather than *de novo*.[8] The distinction may be fine, but is nevertheless important. A *de novo* hearing entirely disregards whether the court below committed any errors, whereas an appeal by way of re-hearing is successful only if the appellate court is satisfied that there is a legal, factual or discretionary error on the part of the court below, having regard to all the evidence in the appeal.[9] The consequence is that the NICA need not have set out its reasoning or rehearsed the factual matrix in such breadth, if it largely agreed with Colton J's findings (which it did).

GROUND 1: THE ACTS OF UNION

Justiciability and timeliness

Before the majority commenced its substantive analysis, two preliminary issues were considered: justiciability and delay. The issue on justiciability arose as one of the appellants' arguments was that the Withdrawal Agreement (WA) was *itself* unlawful by virtue of inconsistency with article VI of the Acts of Union 1800.[10] The majority rejected this on three grounds: first, it conflicted with the well-known rule[11] that international treaties are not justiciable in domestic law unless incorporated;[12] second, the WA was part of a 'distinctly political process', which was not amenable to judicial review;[13] and third, article VI of the Acts of Union did not purport to bind future UK Parliaments.[14] However, this conclusion did not prevent an assessment of the legality of provisions enacting the WA, including the Protocol.

These two statements (treaty-making is non-justiciable and it is still possible to assess the legality of the WA/Protocol) may seem oddly juxtaposed, but the point here is that the *making* of the WA/Protocol in the international plane is non-justiciable, in contrast to the *incorporated* WA/Protocol, which is justiciable.

On delay, the majority agreed with Colton J that the challenge was out of time as it was brought long after three-months post-ratification of the WA,[15] but extended time due to the constitutionally important issues raised.[16]

---

8 Rules of the Court of Judicature in Northern Ireland 1980, Order 59, r 3(1).
9 See eg *Allesch v Maunz* [2000] HCA 40 (Australia), [23].
10 *Allister* (n 7 above) [36].
11 See *R (Miller) v Prime Minister* [2019] UKSC 41, [2020] AC 373, [55]–[56].
12 *Allister* (n 7 above) [37].
13 Ibid [39].
14 Ibid [39].
15 Ibid [50].
16 Ibid [57].

Parliamentary sovereignty

Before addressing the grounds of challenge, it is worth exploring the majority's canvas of the parliamentary sovereignty case law. This was conducted because parliamentary sovereignty was a theme running throughout the claim.

The majority first considered *Thoburn v Sunderland City Council*,[17] which considers parliamentary sovereignty in the context of tensions between 'constitutional' and 'ordinary' statutes. They observed that *Thoburn*'s essential principle was that 'constitutional' statutes cannot be impliedly repealed.[18] The next case was *R (HS2) Action Alliance Limited v Secretary of State for Transport*,[19] concerning an alleged conflict between EU and domestic law (when the UK was still an EU Member State and subject to the doctrine of EU law primacy). Although the Supreme Court found no such conflict in *HS2*, they expressed *obiter* that 'there may be fundamental principles [of constitutional law] which Parliament when it enacted the European Communities Act 1972 (ECA) did not either contemplate or authorise the abrogation'.[20] The next case was *R (Miller) v Brexit Secretary*,[21] concerning whether the UK's exit from the EU required primary legislation to effect. The majority considered that the finding that such legislation was required strongly affirmed parliamentary sovereignty.[22] The next case was the *Treaty Incorporation Bills Reference*[23] concerning whether it was within the Scottish Parliament's competence to enact certain legislation incorporating the United Nations Convention on the Rights of the Child into domestic law. The majority notes the following in the Supreme Court's judgment: 'Parliament can itself qualify its own sovereignty, as it did when it conferred on the courts the power to make declarations of incompatibility … under section 4 of the Human Rights Act'.[24]

The majority concludes from the cases set out thus far that Parliament can limit its own sovereignty, with examples being the ECA 1972 and the Human Rights Act 1998 (HRA); however, no court has ever

17 [2002] EWHC 195 (Admin); [2003] QB 151.
18 *Allister* (n 7 above) [108]–[109].
19 [2014] UKSC 3, [2014] 1 WLR 324.
20 *Allister* (n 7 above) [111].
21 [2017] UKSC 5, [2018] AC 61.
22 *Allister* (n 7 above) [115].
23 [2021] UKSC 42, [2021] 1 WLR 5106.
24 *Allister* (n 7 above) [120]. This is a controversial view, given that, traditionally, the Human Rights Act 1998 has been understood *not* to have qualified parliamentary sovereignty in any way: see Mark Elliott and Nicholas Kilford, 'Devolution in the Supreme Court: legislative supremacy, Parliament's "unqualified" power, and "modifying" the Scotland Act' (*UKCLA* 15 October 2021).

ruled an Act of Parliament 'unconstitutional'.[25] This took them to *R (Jackson) v Attorney General*,[26] where this issue was considered: Lord Steyn opined *obiter* that parliamentary sovereignty is hypothetically circumscribable, being itself a construct of the common law.[27] The majority concludes its canvas of the parliamentary sovereignty case law with the *Continuity Bill Reference*,[28] which concerned another reference considering whether a proposed Bill would be within the Scottish Parliament's legislative competence. The Supreme Court emphasised that notwithstanding devolution, only the UK Parliament wields legal sovereignty; the Scottish Parliament's legislative powers are constrained by the Scotland Act 1998.[29] The majority considered that this principle applied equally to the Northern Ireland Act 1998 (NIA).[30] For his part, McCloskey LJ reached the same conclusion,[31] through an exploration of case law and academic literature.

Setting out this canvas is important because we return to it in our critique of the NICA's reasoning on Ground 1 below.

The Substance of Ground 1

On Ground 1, the appellants argued that Colton J's decision offended constitutional principles by permitting implied repeal of a constitutional statute, namely article VI of the Acts of Union. Article VI famously declares:

> [...] his Majesty's subjects of Great Britain and Ireland shall, from and after the first day of January, one thousand eight hundred and one, be entitled to the same privileges, and be on the same footing as to encouragements and bounties on the like articles, being the growth, produce, or manufacture of either country respectively, and generally in respect of trade and navigation in all ports and places in the united kingdom and its dependencies; and that in all treaties made by his Majesty, his heirs, and successors, with any foreign power, his Majesty's subjects of Ireland shall have the same privileges, and be on the same footing as his Majesty's subjects of Great Britain.[32]

Further, there was an inconsistency between section 7A of the European Union Withdrawal Act 2018 (EUWA), which gave effect to the WA, and article VI of the Acts of Union. The majority broke this down into four questions:

---

25 Ibid [123].
26 [2005] UKHL 56, [2006] 1 AC 262.
27 *Allister* (n 7 above) [124].
28 [2018] UKSC 64, [2019] AC 1022.
29 *Allister* (n 7 above) [128].
30 Ibid [130].
31 Ibid [363].
32 Union with Ireland Act 1800, art VI; Act of Union (Ireland) 1800, art VI.

(i) Is there any inconsistency between the two statutes?
(ii) If so, what is the later statute's effect on the earlier one?
(iii) What was Parliament's intention in enacting the later statute?
(iv) Does the later statute so offend fundamental principles to be rendered unlawful?[33]

For (i), the majority agreed with Colton J that there was some inconsistency: article VI's meaning was clear and unambiguous – all UK citizens were to have the same rights in terms of trade;[34] however, 'in some respects the EUWA ... bring[s] about a difference in treatment'[35] between citizens of Northern Ireland and those in the rest of the UK – specifically, because of the 'additional checks imposed on GB origin goods sent to NI',[36] and 'because the citizens of NI remain subject to some EU regulation and rules' which do not apply to other UK citizens.[37] Turning to (ii), the majority observed that section 7A(3) of the EUWA's scope was very broad, requiring that 'every enactment', which includes Acts of Parliament, must be read subject to the EUWA. The majority considered that this included article VI of the Acts of Union.[38] This did *not* mean, however, that section 7A purported to *repeal* article VI, explicitly or impliedly. In the majority's view, this was because the 'Protocol is not codified as a permanent solution and is drafted in flexible terms'.[39] Rather than being repealed, '[t]he terms of Article VI are subject to the Protocol and so are clearly modified to the extent and for the period during which the Protocol applies'. There was thus no conflict with *Thoburn*, as the issue of implied repeal did not arise.[40] On (iii), notwithstanding their somewhat 'novel' conclusion on (ii) as to whether article VI was repealed, the majority concluded that Parliament's intent in enacting section 7A of the EUWA was clear: there could be no suggestion that Parliament was unaware of the changes brought.[41] On (iv), the 'fundamental' principle considered was the principle of legality.[42] The majority held that it was not engaged, as there was 'no basis' for contending that Parliament had interfered with fundamental rights, whether in the European Court of Human Rights (ECHR) or at common law.[43] The majority thus dismissed

33 Ibid [173].
34 Ibid [183].
35 Ibid [186].
36 Ibid [184].
37 Ibid [185].
38 Ibid [189].
39 Ibid [193].
40 Ibid [195].
41 Ibid [197].
42 *R v Secretary of State for the Home Department ex parte Simms* [1999] UKHL 33, [2000] 2 AC 115.
43 *Allister* (n 7 above) [199].

Ground 1; the summary conclusion being that section 7A of the EUWA was enacted by a sovereign Parliament which knew what the legislation involved.[44]

In his concurring judgment, McCloskey LJ came to a similar conclusion, that article VI was modified, rather than repealed, explicitly or impliedly, by the incorporation of the WA/Protocol.[45]

There are four main points to be made here in analysing the majority's reasoning. First, the length of the majority's excursion (a four-stage enquiry to determine any inconsistency between article VI and the EUWA) was unnecessary, given the respondent's lack of resistance to the idea that the EUWA in fact altered trading arrangements between Great Britain and Northern Ireland, in contrast with the 'same footing' command in article VI.[46] Second, by stating that the Protocol is 'not codified as a permanent solution' with reference to safeguarding measures in article 16 and the consent process in article 18, the majority appears to assume, at least in part, that the use of either provision would alter the customs and regulatory border in some major way (otherwise there would be no need to advert to its temporary nature). This is surprising, given that article 16 measures are required to be 'strictly necessary' and those which 'least disturb' the functioning of the Protocol,[47] rather than providing the means to fundamentally alter, far less dismantle the central feature of the Protocol; and that it is far from clear that an article 18 vote to disapply the EU Single Market provisions on goods would lead to a new borderless reality between Great Britain and Northern Ireland.[48] In pointing to the Protocol's non-permanence, the majority appears to gloss over the context in which the Protocol has to operate, in contrast with its detailed exploration of the context in which the Protocol was incorporated (the parliamentary process). Third, the majority's clear distinction between one statute impliedly repealing another (which 'this case is not about')[49] and one statute modifying the effect of another is another surprise given the majority's earlier exploration of a Supreme Court judgment which *equated* the two concepts.[50] In the

44 Ibid [206].
45 Ibid [392].
46 Ibid [185].
47 Protocol on Ireland/Northern Ireland Agreement on the withdrawal of the United Kingdom of Great Britain and Northern Ireland from the European Union and European Atomic Energy Community [2019] OJ C 384 I/101, art 16.1.
48 S McBride, 'Read the small print – the Irish Sea border may be impossible to remove, even if MLAs vote it down' (*The Newsletter* 30 January 2021) .
49 Allister (n 7 above) [195].
50 Ibid [119].

Treaty Incorporation Bills Reference,[51] the Supreme Court applied its own understanding of the word 'modify' in an earlier judgment – the *Continuity Bill Reference*, in which modification was held to encompass implied repeal.[52] Fourth, the majority's observations on the principle of legality are surprisingly narrow. Legality, as originally defined by Lord Steyn, encompassed the following: 'Parliament legislates for a European liberal democracy founded on the principles and traditions of the common law.'[53] Fundamental rights form part of this liberal democratic paradigm, but at its heart lies the presumption that Parliament does not legislate to radically alter the 'previous policy' of the law without expressing itself clearly.[54] As we set out below, Parliament did indeed clearly express its intentions to radically alter the previous policy of the law, so, contrary to the majority's view in *Allister*, the EUWA *satisfies* the principle of legality.

McCloskey LJ's concurring judgment is also problematic. When exploring constitutional statutes, for example, the judge cited Lord Bingham's well-known (in Northern Ireland, at least) categorisation of the NIA as 'in effect a constitution', and the consequent interpretational approach to that statute, in *Robinson v Northern Ireland Secretary*.[55] However, McCloskey LJ then appears to have drawn a straight line between *Robinson*, Lord Bingham's approach to the Constitution of Belize in *Reyes v The Queen*, and McCloskey LJ's own endorsement of that approach when examining the Constitution of Trinidad and Tobago in *Commissioner of Prisons v Seepersad*.[56] With respect, such an equivalence is unconvincing. Whether or not one agrees with Lord Bingham's characterisation of the NIA as constitutional,[57] it cannot operate to render any other parliamentary statute void – in contrast with the constitutions explored in the other two cases.[58] Thus, the constitutional character of a UK statute is fundamentally different from either the Belizean or Trinidadian Constitutions. Moreover, the factual circumstances in both *Reyes* and *Seepersad* were worlds away from those in *Allister*: *Reyes* involved a constitutional challenge to a mandatory death sentence,[59] while *Seepersad* involved a constitutional

51 *Treaty Incorporation Bills* (n 23 above) [11].
52 *Continuity Bill* (n 28 above) [51].
53 *Regina v Home Secretary, ex parte Pierson* [1998] AC 539, 587C.
54 Ibid 587E.
55 Ibid [336].
56 Ibid [338].
57 See eg the discussion of the Inner House of the Court of Session in *Imperial Tobacco v Lord Advocate* [2012] CSIH 9, [71] (Lord Reed).
58 Constitution of Belize (1981), s 2(1) (constitutional supremacy over ordinary law enacted by the National Assembly of Belize); Constitution of Trinidad and Tobago (1976), s 2 (constitutional supremacy over ordinary law).
59 [2002] UKPC 11 [2002] 2 AC 235, [1].

challenge to judicial detention of minors charged with murder.[60] Indeed, equating *Robinson* with *Reyes* and *Seepersad* might even appear as an argument against having constitutional statutes, if the designation is apt to lead to comparison with constitutional supremacy.

The judge also (like the majority) explored a range of academic literature, concluding that an absolutist reading of parliamentary sovereignty may be qualified by the operation of constitutional statutes,[61] but that the tendency is increasingly *not* to have any such qualifications in a devolved context.[62] As to the first point, McCloskey LJ pointed to a paper by Paul Craig,[63] published before an arguable shift in the judicial understanding of parliamentary sovereignty occurred in *Miller 2/Cherry*.[64] As to the second point, the judge cited an analysis of the *Continuity Bill Reference* by Aileen McHarg and Chris McCorkindale,[65] in which the authors devoted considerable space to exploring the ramifications of the Supreme Court's understanding of how parliamentary legislation is modified (and which encompasses implied repeal).[66] Curiously, this exploration is absent in McCloskey LJ's concurrence, given that this exploration would have given pause to the judge's conclusion (aligning with that of Keegan LCJ) that modification and implied repeal are *distinct* outcomes.

The net effect of the majority and concurring judgments on Ground 1 was, in large part, agreement with Colton J, but with a much wider exploration of academic literature than in the High Court. We would respectfully suggest that such an exploration was both unnecessary and somewhat problematic. Now, we recognise that it is neither appropriate nor necessary for a court to conduct an exhaustive review of academic literature on a point of law before deciding it; indeed, the focus of the judge and that of the academic are fundamentally different.[67] What we aim to offer here is the 'legal reasoning – designed to produce practical

60 [2021] UKPC 13 [2021] 1 WLR 4315, [3].

61 *Allister* (n 7 above) [344].

62 Ibid [363].

63 Ibid [344].

64 *Miller 2* (n 11 above), see Aileen McHarg, 'Giving substance to sovereignty' in Brice Dickson and Conor McCormick (eds), *The Judicial Mind: A Festschrift for Lord Kerr of Tonaghmore* (Hart 2021) 217. Incidentally, insofar as this is a shift in the judicial understanding of parliamentary sovereignty, such a shift has been contextualised by both Aileen McHarg and Jason Varuhas within developing judicial attitudes towards the principle of legality, which the NICA unanimously decided was not engaged in *Allister*. See Jason Varuhas, 'The principle of legality' (2020) 79(3) Cambridge Law Journal 578.

65 *Allister* (n 7 above) [362].

66 Aileen McHarg and Chris McCorkindale, 'The Supreme Court and devolution: the Scottish Continuity Bill reference' (2019) Juridical Review 190, 193–194.

67 Lord Burrows, 'Judges and academics, and the endless road to unattainable perfection' (2022) 55(1) Israel Law Review 50, 54–55.

justice' which '[t]he courts want'.[68] In that spirit, the above critiques are important in a purely doctrinal context. The fact that Craig's paper was published before a shift towards a more substantive understanding of parliamentary sovereignty means that his understanding of the operation of constitutional statutes, and specifically their relationship to parliamentary sovereignty, may not survive the shift. But a doctrinal analysis also reveals a curious omission from both the majority and concurring judgments: any reference to *MacCormick v Lord Advocate*.[69] *MacCormick* is important because it speaks directly to the issue confronting the NICA under the first ground of appeal: namely to determine the interplay between article VI and section 7A of the EUWA. Briefly, *MacCormick* concerned the lawfulness of the monarch being 'Elizabeth II' in Scotland, as Scotland (pre-Union) had never had a monarch styled Elizabeth *I*. The petitioners in *MacCormick* had invoked article I of the Treaty of Union 1707, which had united England (and Wales) and Scotland into one kingdom, submitting that the article precluded there being an Elizabeth II by implication. The petition was dismissed in the Outer House of the Court of Session, *inter alia* because of the unqualified nature of parliamentary sovereignty.[70] In the Inner House, however, the Lord President doubted whether the UK Parliament possessed unqualified sovereignty, being a creature of a treaty, which contained 'unalterable' elements.[71] Nevertheless, the petitioners' reclaiming motion in the Inner House failed because, *inter alia*, there was no authority for the proposition that the courts could scrutinise Acts of Parliament for their compliance with the Treaty of Union.[72] In so far as Parliament incurred any cost of making laws in breach of the Treaty of Union, such cost would be political and not legal: a very narrow view on justiciability, but one which is a necessary implication of there being no authority (before or since *MacCormick* was decided) suggesting any alternative view.[73]

The point of the above discussion is to highlight the proverbial elephant in the (court)room: despite the NICA's admirable focus on discerning *one* answer to the question of what happened to article VI, and on reconciling the political reality of Brexit with the logic of constitutional operation which that court identified, the answer which

68 Ibid 55.

69 1953 SC 396 (Inner House). We do not know whether the case was cited to the NICA.

70 Ibid 403, per Lord Guthrie.

71 Ibid 411–412, per Lord President Cooper.

72 Ibid 412, per Lord President Cooper.

73 See also *Gibson v Lord Advocate* 1975 SC 136 (Outer House) 144, per Lord Keith. Indeed, this issue has some vintage in the Scottish courts, see eg *Laughland v Wansborough Paper Co Ltd* 1921 SLT 341 (Bill Chamber).

resulted from this focus is not entirely persuasive when scrutinised using relevant developments in jurisprudence and academia.

The NICA's approach can be contrasted with that of the High Court, where Colton J's attention was less focused on finding a single answer, and instead more focused on how to give effect to a *present-day* statute. Let us recall that Colton J did not conclusively state whether article VI had been modified, repealed (explicitly or implicitly), rendered obsolete or spent, or indeed anything else. Rather, the first instance judge proceeded on the basis that the two statutes had been enacted in radically different eras, and that 'insofar as there is any conflict between them section 7A [of the EUWA] is to be preferred and given legal effect'.[74] Incidentally, Colton J's approach aligns in some major ways with a comment which Keegan LCJ quoted, and which, we respectfully suggest, would have served the NICA better in arriving at a conclusive answer:

> Certain statutes alter the constitutional arrangements of the United Kingdom in such a way as create a new framework within which later legislation is to be construed and applied. That does not of course preclude a later statute from expressly repealing or amending these new arrangements for it is of the essence of the notion of Parliament's sovereign supremacy that no one Parliament can fetter the scope of action of a later Parliament. But it does mean that the courts will assume—in accordance with the wish of the Parliament enacting the constitutional statute—that no future Parliament intends to depart or contravene any aspect of the new constitutional arrangements unless it does so in clear and unambiguous words.[75]

Periodically, in response to political developments, Parliament enacts statutes which radically alter constitutional arrangements. These statutes mark eras of constitutional operation which differ from others because the arrangements fundamental to constitutional operation differ between different eras. The judicial approach to such constitutional statutes should be to prefer the latter to the former. If the former survives unamended (in any capacity) into the latter's era, it should be held to have been impliedly repeated to the extent necessary to give the latter effect. In this model, an era has a specific context. A constitutional statute speaks to vertical relationships between citizen and state and may also speak to horizontal relationships between different organs of the state.[76] An era therefore is the sum of these relationships and how rights and obligations relating to citizens and state organs operate consistently with the text of a constitutional

74 [2021] NIQB 64, [95]–[114].

75 *Allister* (n 7 above) [113], citing D Greenberg, *Craies on Legislation* 12th edn (Sweet & Maxwell 2021) 14.4.6.

76 Deb (n 6 above) 143.

statute. In the Protocol era, the sweeping 'same footing' command of article VI has been replaced by a customs and regulatory border given domestic effect by the EUWA, with Northern Ireland continuing to be subject to a suite of EU laws beyond simply those relating to goods, none of which apply to Great Britain.

Moreover, where ordinary statutes survive unamended into a new constitutional era, those statutes must then be interpreted consistently[77] with the enacted requirements (as well as the underlying purpose and necessary implications) of the new era. Incidentally, this is what the EUWA requires in any event.[78]

A statute which irresistibly repeals a constitutional statute (or any provision of a constitutional statute) would itself be capable of effecting radical change to constitutional arrangements, thus beginning a new era of constitutional operation. We should stress that 'era' in this context does not necessarily signify the complete end of one constitutional statute and the beginning of another. Rather, it is entirely possible for one or more provisions in a constitutional statute to be repealed (explicitly or impliedly) by a later constitutional statute, so that parts of the previous statute remain in operation, but subject to those newer constitutional provisions which repealed the older ones. A prime example is the Northern Ireland Protocol Bill – by explicitly disapplying large sections of the Protocol, the Bill, if enacted, would in fact radically alter constitutional arrangements in the UK. This is apparent in the far-reaching scope of the powers conferred by the Bill on UK ministers, including the power to give domestic effect to any international agreement replacing, supplementing or modifying the Protocol, by secondary legislation,[79] a role reserved traditionally for Parliament. Nevertheless, the Bill will not (even if enacted as it currently stands) completely undo the constitutional realignment between Great Britain and Northern Ireland brought about by the Protocol. For example, the continued application and evolution of EU equality and non-discrimination law (including the relevant sections of the EU *acquis*) listed under annex I of the Protocol[80] is not 'excluded provision' within the definition of the Bill.[81] Thus, Northern Ireland will continue to be subject to aspects of EU law, both in its current state and as it evolves, despite the radical changes proposed by the Bill. Vertical and horizontal relationships will therefore continue to be different on either side of the Irish Sea, for example with employment

77 In so far as the ordinary statute in question is *pari materia* with a constitutional statute.

78 EUWA 2018, s 7A(3).

79 Northern Ireland Protocol Bill HL Bill (2022–23) 52, cl 19.

80 See also, Protocol (n 47 above) art 2.

81 Northern Ireland Protocol Bill HL Bill (2022–23) 52, cl 14(2).

rights continuing to be subject to certain EU laws which no longer apply in Great Britain, and with Stormont's legislative competence continuing to be subject to certain EU laws,[82] unlike its Scottish and Welsh counterparts.

Applying the above, the operation of the Acts of Union, in their original context (before the partition of Ireland and Irish independence) marked one era of constitutional operation, changing to another with the emergence of the Irish Free State and ultimately the complete severance of constitutional relationships between the UK and Ireland. This explains why decisions such as the *Earl of Antrim's petition*[83] were handed down without a definitive, *single* answer as to what happened to the provisions of the Acts of Union conferring a right upon Irish peers to sit in the House of Lords. A single answer, predicated as it would be on the assumption of unaltered constitutional continuity, was moot: constitutional relationships and operation had been realigned by Parliament so that the earlier era of constitutional operation had, in the context of the Irish peers and their right to sit in the House of Lords, ceased to exist. Exiting the EU realigned relationships and operation again, through the enactment of the EUWA (as amended). In that sense, the NICA was being asked to reconcile different eras of constitutional operation. Insofar as it tried to do as asked, its reasoning consequently suffered. This is why, for example, discussions of legislative intent behind the Acts of Union[84] are out of place (and indeed, era). Whatever the intent behind a previous era, it yields to the intent behind the present one.

We pause here to acknowledge that Mark Elliott has made a version of the argument we have made here,[85] and that our attempts seek to supplement Elliott's.[86] The discussion here is not an attempt at reinventing the wheel, doctrinally speaking. The majority's focus on the legislative intent of the EUWA is, in our view, entirely correct. Rather, this discussion is about setting the NICA's reasoning within wider constitutional doctrine, both as this case proceeds up the appellate hierarchy and more generally, for future litigation of this kind, which may occur with greater frequency given recent developments. Colm Ó Cinnéide charts these, at times, intensely polarised developments in

82 NIA 1998, s 6(2)(ca) (restriction on legislating in breach of art 2(1) of the Protocol, being the rights of individuals).

83 [1967] 1 AC 691 (HL).

84 *Allister* (n 7 above) [377].

85 Mark Elliott, 'Constitutional legislation, European Union law and the nature of the United Kingdom's contemporary constitution' (2014) 10(3) European Constitutional Law Review 379, 387.

86 In addition to supplementing a point about implied repeal of a constitutional statute needing to avoid a constitutional vacuum, made in the comment around the High Court judgment in *Allister*, see Deb (n 6 above) 149.

which rival narratives of constitutional 'fidelity' have attempted self-legitimation through an appeal to different histories of constitutional development, each presented as a continuum which is either being ruptured or reclaimed.[87] And nor is this by any means a recent phenomenon. In his examination of the development of the common law, historian J G A Pocock observed, perhaps with a hint of sardonicism:

> The English supposed that the common law was the only law their land had ever known, and this by itself encouraged them to interpret the past as if it had been governed by the law of their own day; but in addition the fact that the common law was a customary law, and that lawyers defined custom in a way which heavily emphasised its immemorial character, made even more radical the English tendency to read existing law into the remote past.[88]

The recognition that historical laws of a constitutional character must be understood in their time emerges in recent jurisprudence – for example, around the question of the present-day justiciability of the provisions of the Scottish Claim of Right 1689.[89] Indeed, it is possibly the only way to answer the question of how to reconcile two provisions of constitutional character – the recognition that they operate in different *eras*. Such recognition reinforces the sovereignty of the Crown in Parliament, which is ultimately one of the only constant principles[90] which traverse different eras of constitutional operation; not necessarily because of any particular justification grounded in high constitutional theory, but because there is simply *no other choice*. Parliamentary sovereignty may be justified by reference to its democratic credentials,[91] but it has remained undiminished in its potency because the UK recognises no higher source of law to legally constrain Parliament's law-making ability. Those who roam the intellectual wilderness in search of a single, complete and timeless[92] account of the UK constitution dismiss its fundamentally positivist character at their own peril. The very fact that parliamentary

87 Colm Ó Cinnéide, '"You can't go home again": constitutional fidelity and change in post-Brexit Britain' (Public Law Conference, University College Dublin July 2022).

88 J G A Pocock, *The Ancient Constitution and The Feudal Law* (Cambridge University Press 1987) 30–31.

89 *Cherry v Lord Advocate* [2019] CSIH 49, 2020 SC 37, [85], per Lord Brodie.

90 We would also suggest the rule of law as another constant principle.

91 See eg *R (SC and Others) v Work and Pensions Secretary* [2021] UKSC 26, [2022] AC 223, [169]; *R (Bancoult) v Foreign and Commonwealth Secretary* [2008] UKHL 61, [2009] 1 AC 453, [35]; Paul Craig, 'The Supreme Court, prorogation and constitutional principle' (2020) Public Law 248, 254–255; and Jeffrey Goldsworthy, *The Sovereignty of Parliament* (Open University Press 2001) 254.

92 To quote Ó Cinnéide, a 'prelapsarian' account of the UK constitution.

sovereignty appears to be one of the only entrenched[93] principles of this constitution further reinforces its positivist tendencies. The UK constitution is not 'a "thing" to be discovered; it is a term around which competing interests struggle to establish the authority of their own favoured view'.[94] Thus, constitutional realignment is a feature, not a bug, of the UK constitution. Moreover, while this tendency towards constitutional realignment is made possible because a sovereign Parliament is unable (legally) to bind its successors, the inability of one Parliament to bind a future Parliament is reinforced precisely at each moment of constitutional realignment, resulting in realignments as seismic as Brexit. Indeed, a reading of history reveals previous realignments which at least arguably conflicted with article VI.[95]

Recognising different eras of constitutional operation also requires the recognition that, in litigation, courts must determine the necessary implications which flow from and the underlying purpose of constitutional realignment. In eras past, authoritative voices[96] have inferred constitutional purpose from the prevailing religious and socio-political thinking and experience relevant to those eras. Few, however, would argue that, with the modern 'orgy of statute-making',[97] courts need to (or should) reach for purpose in scripture. In line with the positivist tendencies fundamental to the UK constitution, the purpose of constitutional alignment (and realignment) revolves primarily[98] (some would argue, *solely*)[99] around the text of enacted law. We acknowledge that this is not an academically complete account of the constitution but remind ourselves that an academically complete account is unnecessary (and indeed, perhaps unhelpful) for the exercise of the judicial function.

---

93 *Not* timeless – see eg Goldsworthy (n 91 above)159–164 or *Jackson* (n 23 above), [102] per Lord Steyn.

94 Martin Loughlin, 'In search of the constitution' in O Doyle, A McHarg and J Murkens (eds), *The Brexit Challenge for Ireland and the United Kingdom* (Cambridge University Press 2021) 332.

95 C R G Murray and Daniel Wincott, 'Partition by degrees: routine exceptions in border and immigration practice between the UK and Ireland, 1921-1972' (2020) 47 Journal of Law and Society S145.

96 William Blackstone, *Commentaries on the Laws of England* 15th edn (Cadell & Davies 1809) 154, on the tyranny of Charles I's Long Parliament as a lesson in keeping legislative powers in check; John Locke, *An Essay concerning Human Understanding with the Second Treatise of Government* (Wordsworth 2015) 325, on legislative power requiring conformity to 'the will of God'.

97 Grant Gilmore, *The Ages of American Law* (Yale University Press 1977) 95.

98 See eg *Uber BV v Aslam* [2021] UKSC 5, [2021] 4 All ER 209, [70].

99 See eg the differing perspectives in *R(O) v Home Secretary* [2022] UKSC 3, [2022] 2 WLR 343, between Lord Hodge at [30]–[31] and Lady Arden at [65]–[68].

Thus, in our view, the first ground of appeal is misconceived. The framing of article VI, as indeed any other statutory provision, must begin with discerning the will of a sovereign Parliament, and the unimpeachable reality that its will cannot be voided by operation of law, historical or otherwise. This is why, for example, EU membership did not deprive Parliament of its sovereignty, because Parliament retained the ability to repeal the very enactment through which the supremacy of EU law entered the domestic plane.[100]

We stress here that, although we have attempted an answer premised on implied repeal, the operation of the EUWA does not depend on conclusively disposing the question of what has happened to article VI. *Whatever* has happened, as a matter of law, to article VI, the EUWA is to be given effect; to adapt Lord Rodger's phrase (though perhaps not his sentiment behind it),[101] *parlamentum locutum*, *iudicium finitum* – Parliament has spoken, the case is closed.

GROUND 2: THE CONSTITUTIONAL QUESTION IN THE NIA

On Ground 2, the issue was whether the Protocol conflicted with section 1(1) of the NIA, which provides that Northern Ireland remains in the UK unless a majority of its people vote to secede from it. The appellants argued: (i) section 1(1) protected against any substantial change to the Union; (ii) as the Protocol and EUWA effected substantial changes to the Union, a referendum was required; (iii) as no referendum was held, the Protocol and EUWA were unlawful. The majority found this Ground failed at the first hurdle, as section 1(1) only relates to a change in the formal constitutional status of NI. Given the case involved a 'change in intra-UK arrangements brought about by withdrawal from the EU', section 1(1) did not apply to whether the changes enacted by the EUWA and the Protocol were lawful.[102] For essentially the same reasons, McCloskey LJ also concurred.[103]

The NICA's conclusions here should not surprise anyone. We would only add to these conclusions by observing that the principle of consent, as contained in the GFA, speaks not to the 'formal' status of Northern Ireland, as the appellants had characterised it;[104] rather, it speaks to whether the UK or Ireland has *sovereignty* over Northern Ireland. This is not an arid, abstract point; this is a very real matter and provides a complete answer to the appellants' principal challenge here.

100 *Miller 1* (n 21 above) [60].
101 *Home Secretary v AF* (No 3) [2009] UKHL 28, [2010] 2 AC 269, [98].
102 *Allister* (n 7 above) [222].
103 Ibid [409]–[413].
104 Ibid [213].

If a foreign legislature were to make law which is effective in Northern Ireland, such effectiveness results from the exercise of parliamentary sovereignty and not otherwise. That is what Parliament has done in giving effect to the Protocol in domestic law, and any attempt to restrict such effect amounts to considerably qualifying, if not outright negating, the will of Parliament.

GROUND 3: DEMOCRATIC CONSENT IN THE NIA

On Ground 3, the appellants argued that the amendment of section 42 of the NIA by the Protocol on Ireland/Northern Ireland (Democratic Consent Process) EU Exit Regulations 2020 (the 2020 Regulations), removing the cross-community vote process in respect of the democratic consent process in article 18 of the Protocol, was unlawful. They contended that: (i) the democratic consent mechanism in article 18 of the Protocol, as implemented by the 2020 Regulations, was incompatible with the constitutional safeguards (of cross-community voting) in section 42 of the NIA; (ii) the 2020 Regulations were *ultra vires* section 10(1)(a) of the EUWA; and (iii) the 2020 Regulations are inconsistent with the GFA.

There were two main issues to be considered: first, the interpretation of the Henry VIII clauses in the EUWA used to enact the 2020 Regulations, which amended the NIA; and second, the broader constitutional point regarding the democratic consent process. On the first issue, the interpretation of the relevant Henry VIII clauses – in section 8C(1) and (2) of the EUWA – is crucial, and is thus worth setting out in full. Section 8C(1) provides that a minister of the Crown may by regulations make such provision:

> (a) To implement the Protocol in Ireland/Northern Ireland in the Withdrawal Agreement,
>
> (b) To supplement the effect of section 7A in relation to the Protocol, or
>
> (c) Otherwise for the purposes of dealing with matters arising out of, or related to, the Protocol (including matters arising by virtue of section 7A in the Protocol).

Section 8C(2) provides that regulations made under section 8C(1) may amend primary legislation: 'Regulation under sub-section (1) may make provision that could be made by an Act of Parliament (including modifying this Act).'

The appellants argued that the Henry VIII clause should be interpreted restrictively, following *R (Public Law Project) v Lord Chancellor*.[105] The majority concluded that the present claim differed

105 [2016] UKSC 39, [2016] AC 1531.

markedly from the authorities on restrictive interpretation of Henry VIII clauses, as the delegated power here – in section 8C(1) and (2) EUWA – was deliberately conferred in unambiguous terms by Parliament.[106] There was thus no basis for challenging the lawfulness of the delegation.

McCloskey LJ's analysis on this point, although reaching the same conclusion, proceeded slightly differently. Whilst accepting section 8C read as a whole is framed expansively,[107] he considered that this ground of challenge concerns only section 8C(1)(a). In this regard, the operative word 'implement' – which McCloskey LJ considered 'clearly does not extend to variation, repeal, modification or amplification' – was narrow in scope.[108] However, the making of the 2020 Regulations, being 'narrow, targeted and specific', could also be described as 'implementation',[109] which meant that the 2020 Regulations were *intra vires* section 8C(1)(a) EUWA.

On the second issue, the appellants' central argument was that the issue of democratic consent – a central component of the devolution settlement – was impugned by the 2020 Regulations, thus unlawfully offending the 1998 settlement. The majority rejected this for four reasons. First, as stated above, the broad powers in the EUWA gave the Secretary of State the authority to enact the 2020 Regulations. Further, the WA was part of 'international relations', an excepted matter under schedule 2 to the NIA. Second, it was clear that the petition of concern, which engages the cross-community vote process, was only intended for devolved matters. As the WA was not a devolved matter, there could be no argument that section 42 is infringed by article 18 of the Protocol.[110] Third, in relation to the GFA, whilst section 10(1)(a) of the EUWA refers to the need to protect it, there is a difference between a declaration to that effect and justiciable rights under the GFA, which the majority highlight – significantly – is not a part of domestic law.[111] In any event, the form of consent in the article 18 process, even though differing from the NIA's cross-community consent, was a result of considerable political negotiation and subject to parliamentary scrutiny.[112] Fourth, whilst recognising the tension that can arise between devolved legislatures and the UK Parliament in law-making,[113] the NIA permits the Assembly to modify

106 *Allister* (n 7 above) [238].
107 Ibid [424].
108 Ibid [425].
109 Ibid [432].
110 Ibid [243].
111 Ibid [244].
112 Ibid [247].
113 *Continuity Bill Reference* (n 28 above).

provisions only in so far as it is within legislative competence to do so. Given that the conduct of international affairs is not a devolved matter, this Ground had to be dismissed.[114]

We agree only with the majority's first reason for dismissing this ground, namely the breadth of section 8C of the EUWA. This is a sufficient answer to the appellant's challenge, although the amending of primary legislation, especially a constitutional statute (the NIA) by secondary legislation (the 2020 Regulations) is a matter of some concern from the perspective of legislative scrutiny: Parliament cannot amend secondary legislation, but only approve or disapprove of it in its entirety.[115] Nevertheless, section 8C is clear in its scope. The majority's second, third and fourth reasons, however, proceed on the basis that the democratic consent process is an excepted matter. This is somewhat problematic for an important reason: the Secretary of State is responsible for initiating the process, but Stormont is responsible for the vote within that process. Stormont's vote, whether affirming or rejecting articles 5–10 of the Protocol, does not, by itself, affect the conduct of international relations. This is because if articles 5–10 are rejected, it falls on the UK Government and the EU to negotiate their replacement, *not* Stormont. Eliding the role of Stormont within the consent process, with the *reality* of the conduct of international relations, effectively introduces a judicial qualification of excepted matters by the backdoor.[116]

It is true that a vote by Stormont rejecting articles 5–10 is likely to *trigger* a renegotiation between the UK and the EU to replace the relevant articles. However, Stormont's opinion on *what* might replace the articles is not (under the Protocol) a prerequisite to the replacement, so the way in which Stormont's vote *affects* the renegotiation, and thus international relations between the UK and the EU, is not clear cut. It is therefore questionable whether the NICA (and the High Court, for that matter) should have *categorically* classified Stormont's vote as falling within the realm of international relations.

In respect of the second issue under Ground 2, however, the standout aspect of McCloskey LJ's concurring analysis is his observations on the 'juridical identity' of the WA.[117] His essential reasoning is that the

---

114 *Allister* (n 7 above) [249].

115 A point which one of us had also made in the commentary around the High Court judgment in *Allister*, see Deb (n 6 above) 157.

116 The NIA distinguishes between international relations (which is an excepted matter, see NIA, sch 2, para 3) and the observation or implementation of international obligations (which is not an excepted matter, see NIA, sch 2, para 3(c)). For a more detailed version of this argument, see Deb (n 6 above) 155–156.

117 *Allister* (n 7 above) [436].

dual identity of the WA, as both an international treaty and domestic law by primary legislation, permits the reviewing judge to view the 2020 Regulations from an external international perspective, and, in line with the Vienna Convention on the Law of Treaties,[118] it is within the scope of 'international relations' under which the Secretary of State held competence to make the 2020 Regulations.

Here, the judge appears to have taken the WA/Protocol's status as a domestically incorporated international treaty down a somewhat strange path. While it is true that the WA/Protocol is enforceable both in the international and domestic planes, it is only its domestic enforcement which *could* concern the NICA, considering it has no jurisdiction relating to international law. In this context, the judge's conclusion that the WA/Protocol's nature as an international treaty *empowers* the 'making of the 2020 Regulations'[119] is bewildering: nowhere in the text of the treaty is there a power of delegated legislation conferred on the Secretary of State. This conclusion is apparently buttressed by another claim, that article 18.5 of the Protocol 'required domestic legislative action',[120] which is also nowhere to be found in the text of article 18.5. Article 18.2 requires the UK to 'seek democratic consent' in Northern Ireland,[121] in respect of which Parliament authorised the Secretary of State to make secondary legislation (via section 8C of the EUWA) – resulting in the 2020 Regulations. The 2020 Regulations were thus authorised and made *purely* as an exercise of domestic law, in response to a related – though (in important ways) different – obligation arising under international law.

GROUND 4: BREACH OF THE ECHR

On Ground 4, the issue was whether the Protocol violated article 3 of Protocol 1 to the ECHR (A3P1), which protects the right to free elections. This was split into two questions: (i) whether Northern Ireland citizens remaining subject to some aspects of EU law but being unable to vote in European parliamentary elections breaches A3P1; and (ii) whether the differential treatment of Northern Ireland citizens amounts to discrimination contrary to article 14, read with A3P1. On (i), the majority was equivocal as to whether A3P1 was even engaged, giving no firm conclusion, albeit tending towards non-engagement.[122] However, even if it was engaged, the majority considered the interference was

118 Vienna Convention on the Law of Treaties (adopted 23 May 1969, entered into force 27 January 1980) 1155 UNTS 331.

119 *Allister* (n 7 above) [436].

120 Ibid.

121 Protocol (n 47 above) C 384 I/102.

122 Ibid [267].

within the state's wide margin of appreciation on A3P1, taking into account the factors in the Protocol which together encompass a degree of democratic oversight.[123] These included article 18's requirement of democratic consent for ongoing arrangements, article 13.4's provision of a post-enactment information requirement and the requirement that any new act be regulated via the Joint Committee. Further, that Northern Ireland citizens remain enfranchised to vote in both the UK parliamentary and Assembly elections was considered significant.[124]

On (ii), the majority had 'serious reservations' about whether an ancillary test was satisfied, namely the issue of ambit, but proceeded on the basis that it was and applied the four-stage test for determining a breach of article 14: (1) status; (2) differential treatment; (3) lack of reasonable justification; and (4) outside the State's margin of appreciation. On (1), the majority disagreed with Colton J's finding that Northern Ireland residency was a relevant 'status', on the basis that the appellants 'cannot purport to speak for all Northern Ireland residents'.[125] On (2), the majority considered there was no differential treatment between Northern Ireland and other UK residents – no one is able to vote in European parliamentary elections under the Protocol no matter their place of residence.[126] On (3) and (4), the majority considered that the tests were not met, given that there was no group against which differential treatment was to be contrasted with.[127] Both limbs of Ground 4 were thus dismissed.

McCloskey LJ's analysis of this ground proceeded on a largely similar note, albeit giving much greater consideration to the relevant Strasbourg jurisprudence. An interesting observation is McCloskey LJ's highlighting that where it is alleged that there is a lack of proportionality in such matters, it is not uncommon for appellants to propose alternative arrangements which may satisfy all parties: in this case, measures which could facilitate the functioning of the Protocol whilst eliminating the democratic deficit.[128] The absence of any such resolution proposed by the appellants, accompanied by the highly political nature of any such alternate solutions to the Protocol (or lack thereof), supported his view that there is limited scope for judicial intervention on grounds of disproportionality.[129]

---

123 Ibid [268].
124 Ibid [265]–[266].
125 Ibid [278].
126 Ibid [285].
127 Ibid [286].
128 Ibid [484].
129 Ibid.

A notable preliminary point, in our view, is the majority's elision of the distinction between Northern Ireland *citizenship* and *residency*. In laying the groundwork for Ground 4, the two distinct questions identified by the majority referred to Northern Ireland citizens, which differed from the arguments put to the Court by the appellants, which were through the prism of Northern Ireland residency. However, when analysing the article 14 issue, they referred instead to Northern Ireland residency, likely due to the fact much of the relevant Strasbourg jurisprudence refers to residency, as opposed to citizenship.

Moving to the substance, there are two main points in response to the NICA's dismissal of this ground (that A3P1 was engaged but satisfied). First, the appellants were partly correct when stating that the European Parliament continues to act as a legislature for Northern Ireland:[130] *partly*, because it may amend or replace any of the EU legislation mentioned in the Protocol without any further process or scrutiny, and without any limitation on the scope of the amendment or replacement.[131] Neither the UK Government, nor Parliament, nor indeed the Assembly, have any role to play in this process, belying the safety net of democratic accountability which comforted both the majority and McCloskey LJ.[132] Second, the fact that Northern Ireland residents have no right to vote in the European Parliament in respect of this area of law-making (without any further democratic scrutiny) arguably amounts to a blanket ban on the right to vote, and thereby lies outside the UK's margin of appreciation, a point spelled out explicitly in *Hirst v United Kingdom (No 2)*.[133] A more persuasive point might have been to examine whether a right to vote existed at all for Northern Ireland *residents* (and thus whether A3P1 was engaged) following Brexit, given that the right to vote, under EU law, is tied to nationality of a Member State, which the UK has ceased to be.[134]

If A3P1 was not engaged at all, then there would be no need to examine the corresponding article 14 claim, but seeing as the NICA

130 *Allister* (n 7 above) [461].

131 See also, A Deb, 'Parliamentary sovereignty and the Protocol pincer' (2022) Legal Studies 1, 16–18.

132 *Allister* (n 7 above) [267] and [477].

133 (2006) 42 EHRR 41 (GC), [82].

134 Case C-673/20 *EP v Préfet du Gers*, INSEE (CJEU Grand Chamber, delivered 9 June 2022), [58]. We recognise that *EP* was handed down after *Allister but* are unaware whether the NICA were alerted to this fact, and indeed whether the NICA would have delayed its judgment until after *EP* had been handed down. We should also clarify that while we recognise that all Northern Ireland residents are not UK nationals (and indeed include a significant number of Irish, and thus EU, citizens), there is plainly no obligation under EU law for third countries to allow EU citizens resident in their territories to take part in European parliamentary elections.

did so, the majority drew a particularly problematic conclusion around 'other status': the surprising conclusion that the appellants could not claim 'other status' within the meaning of article 14, as the appellants 'cannot purport to speak for all Northern Ireland residents and do not profess to do so'.[135] This conclusion is fundamentally antithetical to the concept of antidiscrimination law. A woman, for example, is not discriminated against as a woman because she can universalise her discriminatory experience to *womankind*.[136] Both the majority and concurring judgments held, however, that 'other status' had to bear some relationship to the listed characteristics in article 14, and that residence was too detached from these characteristics to constitute 'other status'.[137] This conclusion sits somewhat uneasily alongside the fact that residency (albeit in factual circumstances very different from those in *Allister*) has been held to fall within 'other status' in the case law of the Strasbourg Court.[138] However, when looked at in the round, a different picture emerges.

The appellants' claim under article 14 is necessarily tied to their A3P1 claim. A3P1, in turn, provides for a right to vote in the 'choice of the legislature', thus predicated on a law-making body in respect of which a vote may be cast. Now, taking the appellants' claim in this context at its height (that A3P1 is engaged), the European Parliament continues to have the power to make laws – but *only* for Northern Ireland. It has no powers to make law in respect of Great Britain. Northern Ireland residents are therefore in a *relevantly different* situation to those who are resident in Great Britain. Consequently, the article 14 ground falls away. Ultimately, the NICA reaches this exact conclusion.

---

135 *Allister* (n 7 above) [278].

136 See eg *Carvalho Pinto de Sousa Morais v Portugal* [2017] ECHR 719, [52], in which the applicant was held to have been discriminated against as an older woman, in relation to how Portuguese courts had viewed the importance and impact of sexual intimacy in her life, as compared to younger people in general, including younger women.

137 *Allister* (n 7 above) [283] and [536].

138 See eg *Carson v UK* (2010) 51 EHRR 13 (GC), [70]. However, it should be noted that in *Carson* residency was found to be 'other status' on the basis of residency outside the UK state, whereas the appellants in this case have asserted 'other status' based on residence within the different jurisdictions of the UK. It is unfortunate that neither majority nor concurring judgment in this case alluded to, nor engaged with, this point.

GROUND 5: A BREACH OF EU LAW

On Ground 5, the appellants' argued that the Protocol breached two articles of the Treaty on European Union (TEU): (i) article 50, which provides for the process of withdrawal from the EU; and (ii) article 10, which provides that citizens are to be 'represented at Union level in the European Parliament'. On (i), the majority dismissed the argument that article 50(2) did not permit agreement of a future-facing document like the Protocol after withdrawal had taken place, holding that its ordinary and natural meaning permitted specific terms to be set after a process of negotiation/ratification post-withdrawal.[139] On (ii), the majority dismissed the argument on the basis that article 10(2) deals with the functioning of the EU, of which the UK is no longer part.[140] The majority concludes its dismissal of Ground 5 by highlighting that the Ground in fact raised non-justiciable matters, by seeking to impugn the withdrawal process itself, which occurred on the international plane.[141]

McCloskey LJ largely reiterates the majority's decision with respect to the first alleged breach, describing the appellants' contention as 'misconceived'. The judge also draws attention to the 'juridical reality' that, even assuming the UK to have concluded a treaty in breach of the TEU, it is no longer answerable for any such breaches.

Ground 5 gets possibly the shortest treatment from the NICA: both the majority and McCloskey LJ finding that EU law had been complied with in the process of the UK's withdrawal. Neither judgment explores depths of legal reasoning as profound as under Ground 1 (or indeed any other ground), as indeed such exploration is both unnecessary and, more importantly, *pointless*: no UK court is competent to determine the correct interpretation or application of EU law.[142]

CONCLUSION

Allister was as unique in the NICA as it had been in the High Court, but this is hardly surprising. A case which asks seemingly uncomfortable questions about the very foundations of the modern UK and the longstanding orthodoxies undergirding its constitution is a case which comes along but rarely. For that reason alone, the NICA's allowance of an application which began almost a year after the rules of civil procedure allow – perhaps an indulgence in ordinary cases – is not difficult to understand here.

---

139 Ibid [291].
140 Ibid [293].
141 Ibid [293].
142 Not, of course, to be confused with 'retained EU law', which is UK law.

What the case attempts to do is hold a constitutional mirror to the UK Government and Parliament, asking both to account for their decisions and actions in the Brexit saga. But it is precisely because of the nature of the UK constitution, that attempts like these could only be successful in the political arena, which has long since moved onto other issues. Nevertheless, the questions raised in *Allister* await their most authoritative answer yet, as they approach the doors to the Supreme Court.

Articles

Commentaries and Notes